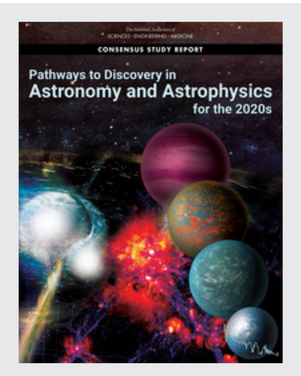


Pathways to Discovery in Astronomy and Astrophysics for the 2020s (2021)

DETAILS

615 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-46734-6 | DOI 10.17226/26141

CONTRIBUTORS

Decadal Survey on Astronomy and Astrophysics 2020 (Astro2020); Space Studies Board; Board on Physics and Astronomy; Division on Engineering and Physical Sciences; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine. 2021. *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26141.

Prepublication Copy – Subject to Further Editorial Correction

Pathways to Discovery in Astronomy and Astrophysics for the 2020s

Committee for a Decadal Survey on Astronomy and Astrophysics 2020 (Astro2020)

Space Studies Board

Board on Physics and Astronomy

Division on Engineering and Physical Sciences

A Consensus Study Report of

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

PREPUBLICATION COPY – SUBJECT TO FURTHER EDITORIAL CORRECTION

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This study is based on work supported by Contract NNH17CB02B with the National Aeronautics and Space Administration, Grant AST-1852611 with the National Science Foundation, Grant DE-SC0019513 with the Department of Energy, and Grant FA9550-19-1-0330 with the Air Force Office of Scientific Research. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any agency or organization that provided support for the project.

International Standard Book Number-13: XXX-X-XXX-XXXXX-X
International Standard Book Number-10: X-XXX-XXXXX-X
Digital Object Identifier: https://doi.org/10.17226/26141

Cover design

Copies of this publication are available free of charge from

Space Studies Board
National Academies of Sciences, Engineering, and Medicine
Keck Center of the National Academies
500 Fifth Street, NW
Washington, DC 20001

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2021. *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26141.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

PREPUBLICATION COPY – SUBJECT TO FURTHER EDITORIAL CORRECTION

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

PREPUBLICATION COPY – SUBJECT TO FURTHER EDITORIAL CORRECTION

COMMITTEE FOR A DECADAL SURVEY ON ASTRONOMY AND ASTROPHYSICS 2020 (ASTRO2020): STEERING COMMITTEE

FIONA A. HARRISON, NAS,[1] California Institute of Technology, *Co-Chair*
ROBERT C. KENNICUTT JR., NAS, University of Arizona and Texas A&M University, *Co-Chair*
JULIANNE DALCANTON, University of Washington
TIM DE ZEEUW, Leiden University
ANDREW S. DRIESMAN, Johns Hopkins University Applied Physics Laboratory
JONATHAN J. FORTNEY, University of California, Santa Cruz
GABRIELA GONZÁLEZ, NAS, Louisiana State University
JORDAN A. GOODMAN, University of Maryland
MARC P. KAMIONKOWSKI, NAS, Johns Hopkins University
BRUCE MACINTOSH, Stanford University
JACOBUS M. OSCHMANN, Independent Consultant
RACHEL A. OSTEN, Space Telescope Science Institute, *Executive Officer*
LYMAN A. PAGE, JR., NAS, Princeton University
ELIOT QUATAERT, NAS, Princeton University
WANDA A. SIGUR, NAE,[2] Independent Consultant
RACHEL SOMERVILLE, Flatiron Institute
KEIVAN G. STASSUN, Vanderbilt University
JEAN L. TURNER, University of California, Los Angeles
PIETER VAN DOKKUM, Yale University
ELLEN G. ZWEIBEL, NAS, University of Wisconsin, Madison

SCIENCE PANELS

PANEL ON COMPACT OBJECTS AND ENERGETIC PHENOMENA

DEEPTO CHAKRABARTY, Massachusetts Institute of Technology, *Chair*
LAURA B. CHOMIUK, Michigan State University
DANIEL E. HOLZ, University of Chicago
RAFFAELLA MARGUTTI, Northwestern University
JULIE McENERY, NASA Goddard Space Flight Center
PETER I. MÉSZÁROS, Pennsylvania State University
RAMESH NARAYAN, NAS,[3] Center for Astrophysics, Harvard and Smithsonian
ELIOT QUATAERT, NAS, Princeton University, *Steering Committee Liaison*
SCOTT M. RANSOM, National Radio Astronomy Observatory
TODD A. THOMPSON, Ohio State University

PANEL ON COSMOLOGY

DANIEL EISENSTEIN, NAS, Harvard University and Sloan Digital Sky Survey III, *Chair*
LINDSEY E. BLEEM, Argonne National Laboratory
MARC P. KAMIONKOWSKI, NAS, Johns Hopkins University, *Steering Committee Liaison*
RACHEL MANDELBAUM, Carnegie Mellon University

---

[1] Member, National Academy of Sciences.
[2] Member, National Academy of Engineering.
[3] Member, National Academy of Sciences.

MIGUEL F. MORALES, University of Washington
DANIEL M. SCOLNIC, Duke University
MATIAS ZALDARRIAGA, NAS, Institute for Advanced Study
KATHRYN M. ZUREK, California Institute of Technology

PANEL ON GALAXIES

DANIELA CALZETTI, NAS, University of Massachusetts, Amherst, *Chair*
MICHAEL BOYLAN-KOLCHIN, University of Texas, Austin
HSIAO-WEN CHEN, University of Chicago
ANN E. HORNSCHEMEIER, NASA Goddard Space Flight Center
SUSAN A. KASSIN, Space Telescope Science Institute
AMANDA A. KEPLEY, National Radio Astronomy Observatory
CHARLES C. STEIDEL, NAS, California Institute of Technology
DANIEL K. STERN, NASA Jet Propulsion Laboratory
TOMMASO TREU, University of California, Los Angeles
PIETER VAN DOKKUM, Yale University, *Steering Committee Liaison*
DAVID H. WEINBERG, Ohio State University

PANEL ON EXOPLANETS, ASTROBIOLOGY, AND THE SOLAR SYSTEM

VICTORIA S. MEADOWS, University of Washington, *Chair*
DAVID A. BRAIN, University of Colorado, Boulder
IAN J.M. CROSSFIELD, Massachusetts Institute of Technology
COURTNEY D. DRESSING, University of California, Berkeley
JONATHAN J. FORTNEY, University of California, Santa Cruz, *Steering Committee Liaison*
TIFFANY KATARIA, Jet Propulsion Laboratory
KATHLEEN E. MANDT, Johns Hopkins University Applied Physics Laboratory
MARK S. MARLEY, NASA Ames Research Center
BRITNEY E. SCHMIDT, Georgia Institute of Technology
CHRISTOPHER C. STARK, Space Telescope Science Institute

PANEL ON THE INTERSTELLAR MEDIUM AND STAR PLANET FORMATION

LEE W. HARTMANN, University of Michigan, *Chair*
SEAN M. ANDREWS, Smithsonian Astrophysical Observatory
PHILIP J. ARMITAGE, Stony Brook University
BRUCE T. DRAINE, NAS, Princeton University
KAITLIN M. KRATTER, University of Arizona
KARIN M. SANDSTROM, University of California, San Diego
SNEZANA STANIMIROVIC, University of Wisconsin, Madison
ELLEN G. ZWEIBEL, NAS, University of Wisconsin, Madison, *Steering Committee Liaison*

PANEL ON STARS, THE SUN, AND STELLAR POPULATIONS

SARBANI BASU, Yale University, *Chair*
NANCY S. BRICKHOUSE, Center for Astrophysics, Harvard and Smithsonian
ADAM BURGASSER, University of California, San Diego
JULIANNE DALCANTON, University of Washington, *Steering Committee Liaison*
JENNIFER A. JOHNSON, Ohio State University
R.T. JAMES MCATEER, New Mexico State University

ELISA V. QUINTANA, NASA Goddard Space Flight Center
LOUIS-GREGORY STROLGER, Space Telescope Science Institute

PROGRAM PANELS

PANEL ON AN ENABLING FOUNDATION FOR RESEARCH

DAVID N. SPERGEL, NAS, Princeton University, *Chair*
MICHAEL BLANTON, New York University
KELLE L. CRUZ, Hunter College
MARK J. DEVLIN, University of Pennsylvania
MEGAN E. DONAHUE, Michigan State University
KEITH A. HAWKINS, University of Texas, Austin
ALINA A. KIESSLING, Jet Propulsion Laboratory
KARIN ÖBERG, Harvard University
ANGELA V. OLINTO, NAS, University of Chicago
BERNARD J. RAUSCHER, NASA Goddard Space Flight Center
RACHEL SOMERVILLE, Center for Computational Astrophysics, Flatiron Institute, *Steering Committee Liaison*
JAMES M. STONE, Institute for Advanced Study

PANEL ON ELECTROMAGNETIC OBSERVATIONS FROM SPACE 1

MARCIA J. RIEKE, NAS, University of Arizona, *Chair*
RUSLAN BELIKOV, NASA Ames Research
REBECCA A. BERNSTEIN, Carnegie Institution for Science
LESTER M. COHEN, Center for Astrophysics | Harvard & Smithsonian
NIKOLE K. LEWIS, Cornell University
BRUCE MACINTOSH, Stanford University, *Steering Committee Liaison*
AMY MAINZER, University of Arizona
MARK P. SAUNDERS, Independent Consultant
EVGENYA L. SHKOLNIK, Arizona State University
GEORGE SONNEBORN, NASA Goddard Space Flight Center
C. MEGAN URRY, NAS, Yale University

PANEL ON ELECTROMAGNETIC OBSERVATIONS FROM SPACE 2

STEVEN M. KAHN, Stanford University, *Chair*
LISA BARSOTTI, Massachusetts Institute of Technology
ALLISON BARTO, Ball Aerospace and Technologies Corporation
MICHAEL BAY, Bay Engineering Innovations
MARTIN ELVIS, Center for Astrophysics | Harvard & Smithsonian
CHARLES J. HAILEY, Columbia University
CHRYSSA KOUVELIOTOU, NAS, George Washington University
CHARLES R. LAWRENCE, Jet Propulsion Laboratory
S. HARVEY MOSELEY, JR., Quantum Circuits, Inc.
RESHMI MUKHERJEE, Barnard College
LYMAN A. PAGE, JR., NAS, Princeton University, *Steering Committee Liaison*
GORDON J. STACEY, Cornell University

PANEL ON OPTICAL AND INFRARED OBSERVATIONS FROM THE GROUND

TIMOTHY M. HECKMAN, NAS, Johns Hopkins University, *Chair*
DAVID A. BEARDEN, Jet Propulsion Laboratory
DAVID CHARBONNEAU, NAS, Harvard University
SUVI GEZARI, University of Maryland
ANDREA M. GHEZ, NAS, University of California, Los Angeles
JENNY E. GREENE, Princeton University
J. TODD HOEKSEMA, Stanford University
JACOBUS M. OSCHMANN, Independent Consultant, *Steering Committee Liaison*
RICHARD W. POGGE, Ohio State University
MASSIMO ROBBERTO, Space Telescope Science Institute
NATASCHA M. FÖRSTER SCHREIBER, Max Planck Institute for Extraterrestrial Physics
DAVID R. SILVA, University of Texas, San Antonio

PANEL ON PARTICLE ASTROPHYSICS AND GRAVITATION

JOHN F. BEACOM, Ohio State University, *Co-Chair*
LAURA CADONATI, Georgia Institute of Technology, *Co-Chair*
DAVID Z. BESSON, University of Kansas
GABRIELA GONZÁLEZ, NAS, Louisiana State University, *Steering Committee Liaison*
JORDAN A. GOODMAN, University of Maryland, *Steering Committee Liaison*
ELIZABETH A. HAYS, NASA Goddard Space Flight Center
N. JEREMY KASDIN, University of San Francisco
DAVID B. KIEDA, University of Utah
ANDREA N. LOMMEN, Haverford College
BRIAN D. METZGER, Columbia University
JAMES H. YECK, University of Wisconsin, Madison
NICOLAS YUNES, University of Illinois, Urbana-Champaign

PANEL ON RADIO, MILLIMETER, AND SUBMILLIMETER OBSERVATIONS FROM THE GROUND

ANDREW J. BAKER, Rutgers, The State University of New Jersey, *Chair*
JACQUELINE N. HEWITT,[4] Massachusetts Institute of Technology, *Vice Chair*
HÉCTOR G. ARCE, Yale University
RAVINDER S. BHATIA, Thirty Meter Telescope
TRACY E. CLARKE, Naval Research Laboratory
MATT A. DOBBS, McGill University
DAVID L. KAPLAN, University of Wisconsin, Milwaukee
DANIEL P. MARRONE, University of Arizona
LYNN D. MATTHEWS, Massachusetts Institute of Technology Haystack Observatory
JOAN R. NAJITA, NSF National Optical-Infrared Astronomy Research Laboratory (NOIRLab)
RICHARD L. PLAMBECK, University of California, Berkeley
JEAN TURNER, University of California, Los Angeles, *Steering Committee Liaison*

[4] Resigned from the panel on April 16, 2020, and did not participate in the panel's deliberations or the writing of its report.

PANEL ON STATE OF THE PROFESSION AND SOCIETAL IMPACTS

MARGARET M. HANSON, University of Cincinnati, *Co-Chair*
ENRICO RAMIREZ-RUIZ, University of California, Santa Cruz, *Co-Chair*
GURTINA BESLA, University of Arizona
PATRICIA T. BOYD, NASA Goddard Space Flight Center
KATHRYNE J. DANIEL, Bryn Mawr College
MARTHA P. HAYNES, NAS, Cornell University
JEDIDAH C. ISLER, Dartmouth University
RACHEL L. IVIE, American Institute of Physics
KATHRYN V. JOHNSTON, Columbia University
CASEY W. MILLER, Rochester Institute of Technology
JESÚS PANDO, DePaul University
JULIE R. POSSELT, University of Southern California
JANE R. RIGBY, NASA Goddard Space Flight Center
WILLIE S. ROCKWARD, Morgan State University
KEIVAN G. STASSUN, Vanderbilt University, *Steering Committee Liaison*

DECADAL SURVEY ON ASTRONOMY AND ASTROPHYSICS 2020 (ASTRO2020) STAFF

COLLEEN N. HARTMAN, Director, Space Studies Board and Board on Physics and Astronomy[1]
JAMES LANCASTER,[2] Director, Board on Physics and Astronomy
GREGORY MACK, Senior Program Officer, Board on Physics and Astronomy, *Steering Committee Co-Study Director and Panel on Compact Objects and Energetic Phenomena*
ABIGAIL A. SHEFFER, Senior Program Officer, Space Studies Board, *Steering Committee Co-Study Director and Panel on Electromagnetic Observations from Space 2*
DIONNA WISE, Program Coordinator, Space Studies Board, *Steering Committee and Panel on State of the Profession and Societal Impacts*
MIA BROWN, Research Associate, Space Studies Board, *Steering Committee*

ALAN ANGLEMAN, Senior Program Officer, Aeronautics and Space Engineering Board, *Panel on Particle Astrophysics and Gravitation*
SHENAE BRADLEY, Administrative Assistant, Computer Science and Telecommunications Board, *Panel on Galaxies*
MEGAN CHAMBERLAIN, Senior Program Assistant, Space Studies Board, *Panel on Electromagnetic Observations from Space 2, Panel on Optical and Infrared Observations from the Ground, and Panel on Particle Astrophysics and Gravitation*
ARTHUR CHARO, Senior Program Officer, Space Studies Board, *Panel on Radio, Millimeter, and Submillimeter Observations from the Ground*
DWAYNE DAY, Senior Program Officer, Aeronautics and Space Engineering Board, *Panel on Electromagnetic Observations from Space 1*
GREG EYRING, Senior Program Officer, National Materials and Manufacturing Board, *Panel on the Interstellar Medium and Star Planet Formation, Panel on an Enabling Foundation for Research, and Panel on State of the Profession and Societal Impacts*
SANDRA GRAHAM, Senior Program Officer, Space Studies Board, *Panel on Stars, the Sun, and Stellar Populations*
GAYBRIELLE HOLBERT, Program Assistant, Space Studies Board, *Panel on Stars, the Sun, and Stellar Populations*
AMISHA JINANDRA, Associate Program Officer, Board on Physics and Astronomy *Panel on Exoplanets, Astrobiology, and the Solar System*
CHRISTOPHER JONES, Program Officer, Board on Physics and Astronomy, *Panel on Exoplanets, Astrobiology, and the Solar System*
JAMES MYSKA, Program Officer, National Materials and Manufacturing Board, *Panel on Cosmology*
DANIEL NAGASAWA, Program Officer, Space Studies Board, *Panel on State of the Profession and Societal Impacts*
ERIK SAARI, Administrative Assistant, Board on Army Research Development (until November 30, 2019), *Panel on Compact Objects and Energetic Phenomena, and Panel on Exoplanets, Astrobiology, and the Solar System*
ERIK SVEDBERG, Senior Program Officer, National Materials and Manufacturing Board, *Panel on Galaxies, and Panel on Optical and Infrared Observations from the Ground*
LINDA WALKER, Program Coordinator, Board on Physics and Astronomy *Panel on Cosmology, Panel On The Interstellar Medium And Star Planet Formation, Panel on an Enabling Foundation for Research, Panel on Electromagnetic Observations from Space 1, and Panel on Radio, Millimeter, and Submillimeter Observations from the Ground*

[1] Since May 2021.
[2] Retired May 2021.

*We want to thank the following—Senior Scholar and former BPA Board Director, Donald C. Shapero; Christine Mirzayan Science and Technology Policy Fellows Sara Crandall and Emily Moravec; and Lloyd V. Berkner Space Policy Interns Robert Bullard, Benjamin Cassese, Katherine Dzurilla, Tarini Konchady, Sarah Moran, Osase Omoruyi, Genevieve Payne, Cindy Vo.

BOARD ON PHYSICS AND ASTRONOMY

ANDREW J. LANKFORD, University of California, Irvine, *Chair*
MEIGAN ARONSON, University of British Columbia
WILLIAM BIALEK, NAS, Princeton University
JILL DAHLBURG, Naval Research Laboratory, ret.
SALLY DAWSON, Brookhaven National Laboratory
LOUIS F. DIMAURO, The Ohio State University
WENDELL T. HILL III, University of Maryland
ALAN J. HURD, Los Alamos National Laboratory
CHUNG-PEI MA, University of California, Berkeley
ANGELA VILLELA OLINTO, The University of Chicago
DAVID H. REITZE, California Institute of Technology
SUNIL SINHA, University of California, San Diego
RISA H. WECHSLER, Stanford University
WILLIAM A. ZAJC, Columbia University

STAFF

JAMES C. LANCASTER,[3] Director
COLLEEN N. HARTMAN,[4] Director
NEERAJ P. GORKHALY, Associate Program Officer
AMISHA JINANDRA, Associate Program Officer
CHRISTOPHER J. JONES, Program Officer
RADAKA LIGHTFOOT, Finance Business Partner
GREGORY MACK, Senior Program Officer
LINDA WALKER, Program Coordinator

[3] Retired May 2021.
[4] Since May 2021.

SPACE STUDIES BOARD

STAFF

Preface

The summary charge for the Committee for a Decadal Survey of Astronomy and Astrophysics (Astro2020; hereafter the "steering committee") reads as follows:

> The National Academies of Sciences, Engineering, and Medicine shall convene an ad hoc survey committee and supporting study panels to carry out a decadal survey in astronomy and astrophysics. The study will generate consensus recommendations to implement a comprehensive strategy and vision for a decade of transformative science at the frontiers of astronomy and astrophysics.

The complete statement of task and its scope is provided in Appendix A of this report. The steering committee, with inputs from 13 expert panels encompassing the breadth of astronomy and astrophysics, was specifically asked to (1) provide an overview of the current state of astronomy and astrophysics science, and technology research in support of that science, with connections to other scientific areas where appropriate; (2) identify the most compelling science challenges and frontiers in astronomy and astrophysics, which shall motivate the committee's strategy for the future; (3) develop a comprehensive research strategy to advance the frontiers of astronomy and astrophysics for the period 2022-2032 that will include identifying, recommending, and ranking the highest-priority research activities; (4) utilize and recommend decision rules, where appropriate, that can accommodate significant but reasonable deviations in the projected budget or changes in urgency precipitated by new discoveries or unanticipated competitive activities; (5) assess the state of the profession, including workforce and demographic issues in the field, identify areas of concern and importance to the community, and where possible, provide specific, actionable, and practical recommendations to the agencies and community to address these areas.

Astro2020 was sponsored by the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), the Department of Energy (DOE) Office of High Energy Physics, and the Air Force Office of Space Research (AFOSR). These federal agencies all participate in different aspects of the U.S. space- and ground-based astronomy and astrophysics program. Internally, the decadal survey effort at the National Academies was a joint project of the Board on Physics and Astronomy and the Space Studies Board. This is the seventh decadal plan for astronomy and astrophysics conducted over the history of the National Academies.

The scope of the science assessed by the decadal survey is broad, encompassing all aspects of observational, theoretical, and computational astronomy including ground-based solar observations, but specific activity recommendations were limited to those administered by NSF Division of Astronomical Sciences and the NASA Astrophysics Division. Scientific areas in astronomy and astrophysics pursued by the DOE Office of High Energy Physics were also included in the study, but activity recommendations were limited to NSF and NASA as described above.

The committee was also tasked with assessing three space projects, WFIRST (since renamed the Nancy Grace Roman Space Telescope), Athena (Advanced Telescope for High-Energy Astrophysics), and LISA (Laser Interferometer Space Antenna)—the latter two being European-led missions with significant NASA participation. These three projects were highly ranked priorities for the 2010 *New Worlds New Horizons* (NWNH) decadal survey, also called Astro2010, and are under development but not yet launched.[1] The committee was invited to comment on the status and future direction of NASA

---

[1] National Research Council, 2010, *New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/12951.

support for these projects, but they were not considered together with nor ranked against new projects proposed for the coming decade. Likewise the survey was invited to consider the status and evolution of ongoing programs within NASA's program of record.

STUDY PROCESS AND PARTICIPANTS

Organization of the Astro2020 decadal survey, including the steering committee and 13 expert panels, began with the appointment of the co-chairs in late 2018. The steering committee consisted of 20 members, including the co-chairs. Members were selected to cover as fully as possible the scientific scope of the survey, the range of observational (ground, space, and particle/gravitational astrophysics) and theoretical disciplines, as well as technical and managerial background in space and/or ground-based facilities, and to comprise as representative a group of experts as possible in terms of individual, institutional, and geographical demographics. National Academies policies governing potential conflicts of interest by steering committee and panel members were strictly enforced. In particular, broad and open-minded thinkers were sought out as opposed to advocates for individual missions or subfields. The role of the Executive Officer was also redefined for this survey, to that of a full voting member of the steering committee, with additional administrative and coordination responsibilities as part of the leadership team.

All meetings and deliberations for the survey operated under a code of conduct to guide discussions. This code of conduct was developed in addition to the National Academies policy on sexual harassment and bullying for committee members, panelists, and meeting attendees, and in addition to a code of conduct for National Academy of Sciences members. The statement includes a recognition of strongly held, possibly differing opinions; a dedication to open dialogue and open exchange of ideas; and professional, civil, and collegial treatment of colleagues so that an open exchange of ideas can occur.

During the course of the panel and steering committee meetings, a number of public information-gathering sessions were held. These included briefings by the agencies (NASA, NSF, DOE), invited presentations by selected projects to the program panels, and other informational sessions sponsored by the other panels or the steering committee. Throughout the survey, the representatives of all three agencies urged the committee to be "ambitious"[2] and "aspirational,"[3] and the committee hopes that it has risen to this challenge.

The charges of the expert panels were similar to that of the Astro2010 survey, but with a few important changes. The number of science panels was increased from five to six panels, to accommodate the very rapid growth over the previous decade of exoplanetary astronomy and multi-messenger astrophysics, while preserving balance across all subject areas. The basic programmatic panel structure from Astro2010 was also retained, except that the Panel on Electromagnetic Observations from Space was divided into two panels, in order to accommodate the very large number of project proposals and community white papers in that area. Finally, two new panels for Astro2020 were appointed—the Panel on an Enabling Foundation for Research (program panel) was charged with evaluating cross-cutting supporting programs (e.g., grants programs, laboratory astrophysics, data archiving and data science, computation, theory), and the Panel on State of the Profession and Societal Impacts (SoPSI) was charged to advise the steering committee on this specific area in its statement of task. These 13 appointed panels (six science, six program, and one state of the profession) comprised 127 members. In addition, a steering committee "liaison" member was appointed to each panel to facilitate the flow of information and communication between panels and the main survey committee. The liaisons participated in the panel discussions but did not hold formal voting rights. In all, 141 individuals participated in the panel deliberations with a purpose of providing input to the steering committee. Each panel drafted its own report, with suggestions for the steering committee to consider, as it held its own deliberations to reach its

[2] Presentation to Astro2020 committee by Paul Hertz, NASA Astrophysics Division, July 2019.
[3] Presentation to Astro2020 committee by Ralph Gaume, NSF Division of Astronomical Sciences, July 2019.

recommendations for the main report. To underscore the importance of the panel reports, they have been published together with this main report as appendixes.

The science panels were asked to provide a brief review of the current state of the science in their topic areas and determine four important science questions to be addressed in the next decade and one area that shows great promise for discovery. The program panels were charged to assess the ability of current and proposed projects under consideration to address the science panels' questions and discovery areas, to comment on the Technical, Risk, and Cost Evaluations (TRACE) of the proposed projects, to identify key areas of technical development or precursor research activities, and to discuss the balance of small, competitively selected activities versus larger strategic investments needed to address the science questions. The program panels were not asked to prioritize or rank projects, but rather to suggest to the steering committee the projects with the best potential to realize the capabilities needed to address the science panels' questions and discovery areas. The Panel on the State of the Profession and Societal Impacts was asked to gather information on the health and demographics of the astronomy and astrophysics community and make actionable suggestions to the steering committee on the topics of demographics, diversity and inclusion, workplace climate, workforce development, education, public outreach, and relevant areas of astronomy and public policy. Further information about the charges to the panels is found in Appendix A.

The information-gathering and deliberative phases of Astro2020 were carefully coordinated. Members of the astronomical community were invited to submit white papers to the survey, and these papers formed the foundation and starting point for all of the panel deliberations. In the first phase, 572 science white papers were received in early 2019. A second call for "activity, project, and state of profession consideration" (APC) white papers in July 2019 elicited 294 responses. Every white paper was assigned to and read by one or more of the panels.

The panel meetings themselves were phased. The science panels each held two formal meetings, the program panels each held three formal meetings, and all held several additional teleconferences. Meetings of the science panels took place during the second half of 2019, so that the priorities emerging from the reports of those panels could be incorporated into the program panel deliberations. The science panel chairs presented their findings to the steering committee and the program panels at a face-to-face meeting in December 2019 and delivered their written reports in early 2020. The program panels' meetings began in November 2019, and they presented their results to the steering committee in May 2020 and delivered initial written reports in June 2020. The SoPSI panel met and deliberated on an independent schedule, including holding a public listening session at the American Astronomical Society meeting on January 6, 2020. The SoPSI report was fully incorporated into the overall deliberations and prioritization phases of the steering committee activities.

During the course of the panel deliberations, a number of other inputs were received, and these were especially important for the program panels. After an initial review of all projects proposed for a given panel area, the panels issued requests for information (RFIs) from selected projects to obtain more detailed information that was initially provided in the respective APCs. These included all of the large space and ground "flagship" proposals and selected examples of smaller projects. Selected projects were also invited to present summaries to their respective program panels in public sessions. Many of these projects then underwent a detailed TRACE study, conducted by an independent contractor (The Aerospace Corporation). This independent analysis was mandated by the 2008 NASA Authorization Act, which "directs the Administrator to enter into agreements periodically with the National Academies for decadal surveys to take stock of the status and opportunities for Earth and space science discipline fields and aeronautics research and to recommend priorities for research and programmatic areas over the next decade."[4] Additionally, the act "requires that such agreements include independent estimates of life cycle costs and technical readiness of missions assessed in the surveys whenever possible." In-house analyses of technology readiness, risk, and cost estimates provided by the project teams themselves supplemented

[4] National Aeronautics and Space Administration Authorization Act of 2008, P.L. 110-422, Section 1104 (October 15, 2008).

this analysis. Details of the TRACE process are provided in Appendix O of this report. This process was formerly labeled “Cost and Technical Evaluation” (CATE) and was conducted for recent National Academies surveys in planetary science and solar and space physics, as well as Astro2010.

The schedule for this review was impacted by two outside events—a 35-day government shutdown from December 2018 to January 2019, and the COVID-19 pandemic of 2020 to 2021. The shutdown happened just as science white papers were being solicited, so the deadline for submissions was delayed by a month. The impacts of the COVID-19 pandemic were much more severe. The initial disruptions in March and April 2020 occurred when the program panels were completing their final meetings. Final panel deliberations were held virtually, and delivery of the panel reports to the steering committee were delayed by up to 2 months as everyone adjusted to the new reality of working, caring for children, teaching, and performing service to the community, all while under a stay-at-home order. The greatest impact was on the deliberations of the steering committee, which needed to replace its remaining schedule of four 3- to 4-day face-to-face meetings (out of six total) with more than 20 all-day Zoom meetings. Early into the pandemic, the survey co-chairs and National Academies’ staff decided not to allow the disruptions to compromise the quality or integrity of the survey, and the inevitable result was a several month delay from the original schedule. Included in these virtual meetings were presentations of preliminary results by the program and SoPSI panels during the summer of 2020.

After the panel reports were received and assembled, the steering committee proceeded with the main prioritization discussions, fully informed by the panel reports. The steering committee addressed a few additional topics that were not taken up in full by a program panel (e.g., satellite constellations and radio frequency interference). In such cases, working groups were appointed within the steering committee or by committee and cross-panel working groups. The steering committee’s deliberations were aided by the introduction of innovative strategies to assist in reaching consensus in the virtual environment necessitated by COVID-19, such as online voting tools, collaborative online document editing, the utilization of various videoconferencing features, and asynchronous deliberations (Figure P.1).

FIGURE P.1 Steering committee members and staff met virtually on May 27, 2021.

Acknowledgments

The completion of an undertaking of this scale and importance relies on the contributions of many, and we conclude this introduction by thanking some of the key people. The members of the steering committee, the panel chairs, and the members of the panels contributed tirelessly to the effort, even after the onset of the pandemic extended the timetables and the complexity of their activities. One individual in particular, Executive Officer Rachel Osten, provided extraordinary service as a logistical and scientific coordinator, valued committee member, and advisor to the co-chairs, and is amply deserving of special mention.

We likewise received unflagging support from the National Academics of Sciences, Engineering, and Medicine, including staff from the Board on Physics and Astronomy and the Space Studies Board who supported the work of the panels and steering committee and facilitated the smooth flow of work, meetings, reports, and other activities. We also gratefully acknowledge the contributions of the National Academies staff, interns, and fellows who participated in the survey at various stages——panel staff Alan Angleman, Shenae Bradley, Megan Chamberlain, Arthur Charo, Dwayne Day, Greg Eyring, Sandra Graham, Gaybrielle Holbert, Amisha Jinandra, Christopher Jones, James Myska, Daniel Nagasawa, Erik Saari, Erik Svedberg, and Linda Walker; Christine Mirzayan Science and Technology Policy Fellows Sara Crandall and Emily Moravec; and Lloyd V. Berkner Space Policy Interns Robert Bullard, Benjamin Cassese, Katherine Dzurilla, Tarini Konchady, Sarah Moran, Osase Omoruyi, Genevieve Payne, and Cindy Vo. Above all we wish to thank the survey study co-directors, Abigail Sheffer and Gregory Mack, and program coordinator, Dionna Wise, for their steadfast efforts in overseeing and managing the survey process. Senior advice and leadership was provided throughout the survey by Colleen N. Hartman, director of the Space Studies Board, and James Lancaster, director of the Board on Physics and Astronomy, who are also deserving of special thanks.

The foundation materials for Astro2020 were the 867 science and activity, project, and state of profession consideration white papers, which were contributed by thousands of authors from around the astronomical community. The quality of those papers testified to the effort and thought that went into their preparation, and we express our heartfelt thanks to everyone who participated in this effort and made this a truly community-based survey.

We are also grateful to the members of our sponsoring agencies, especially Ralph Gaume at the National Science Foundation, Paul Hertz at NASA, and Kathleen Turner at the Department of Energy, and their staffs for sharing their time during the early phases of the survey to provide comprehensive background information and guidance, and for responding to our many requests for further background and statistical information.

The Technical, Risk, and Cost Evaluations assessments of the large projects considered by Astro2020 were carried out by the Aerospace Corporation, and we thankfully acknowledge the contributions of their team members, including in particular Randy Persinger and Matthew Marshall. We also thank the many experts who generously contributed their help to the steering committee or its panels as consultants. These include Steven Battel (Battel Engineering), who provided early advice on the setup of the survey processes, and Justin Jonas (South African Radio Astronomy Observatory), Joseph McMullin (Square Kilometre Array Organisation), James Moran (Harvard-Smithsonian Center for Astrophysics), and Antony Schinckel (Commonwealth Scientific and Industrial Research Organisation), who provided expert technical advice to the Panel on Radio, Millimeter, and Submillimeter Observations from the Ground. Sherard Robbins (University of Arizona) kindly served as the facilitator of the Panel on State of the Profession and Societal Impacts listening session at the American Astronomical Society meeting.

The scientific vision and recommended strategic program arising from this process are presented in the remainder of this report and will not be summarized here. As with all of the preceding astronomy and astrophysics surveys, difficult choices were necessary, but that is the main reason for conducting these decadal reviews.

We hope that we have provided not only an ambitious, inspirational, and aspirational vision and roadmap for the coming decade, but also a pathway towards realizing even greater objectives in the future.

Fiona Harrison and Robert Kennicutt, *Co-Chairs*
Committee for a Decadal Survey on Astronomy and Astrophysics 2020

Acknowledgment of Reviewers

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their review of this report:

Main Report

Stephon Alexander, Brown University
Beverly K. Berger, Retired
Jeremiah K. Darling, University of Colorado, Boulder
Catherine Espaillat, Boston University
Lennard Fisk, NAS,[1] University of Michigan
Kathryn Flanagan, Space Telescope Science Institute
Thomas Greene, NASA Ames Research Center
Jeffrey Hall, Lowell Observatory
Victoria Kaspi, NAS, McGill University
Lisa Kewley, Australian National University
Christopher Kochanek, Ohio State University
Roy Maizel, NASA Headquarters (retired)
Christopher F. McKee, NAS, University of California, Berkeley
Sally Oey, University of Michigan
Scott Tremaine, NAS, Canadian Institute for Theoretical Astrophysics and the Institute for Advanced Study
Michael S. Turner, NAS, Kavli Foundation
Rainer Weiss, NAS, Massachusetts Institute of Technology

Science Panel Appendixes

Bangalore Sathyaprakash, University of Pennsylvania
Catherine Espaillat, Boston University
Katherine Freese, NAS, University of Texas, Austin
Lisa Kaltenegger, Cornell University
Chip Kobulnicky, University of Wyoming
Christopher F. McKee, NAS, University of California, Berkeley
M. Claudia Scarlatta, University of Minnesota
Risa Wechsler, Stanford University

[1] Member, National Academy of Sciences.

Program Panel Appendixes
Panel on An Enabling Foundation for Research
Marcel Agüeros, Columbia University
Beverly K. Berger (retired)
Christopher F. McKee, NAS, University of California, Berkeley
Arfon Smith, Journal of Open Source Software and GitHub

Panel on Electromagnetic Observations from Space 1
Orlando Figueroa, Orlando Leadership Enterprise, LLC
Christopher F. McKee, NAS, University of California, Berkeley
Lisa A. Poyneer, Lawrence Livermore National Laboratory
Dmitry Savransky, Cornell University
Alycia J. Weinberger, Carnegie Institution of Washington

Panel on Electromagnetic Observations from Space 2
Orlando Figueroa, Orlando Leadership Enterprise, LLC
Robert Gehrz, University of Minnesota
Dan McCammon, University of Wisconsin, Madison
Christopher F. McKee, NAS, University of California, Berkeley

Panel on Optical and Infrared Observations from the Ground
Lisa Kewley, Australian National University
K.D. Leka, NorthWest Research Associates
Christopher F. McKee, NAS, University of California, Berkeley
Christy Tremonti, University of Wisconsin, Madison
Robert E. Williams, University of California, Santa Cruz

Panel on Particle Astrophysics and Gravitation
Beverly K. Berger (retired)
Christopher F. McKee, NAS, University of California, Berkeley
Kohta Murase, Pennsylvania State University
Steven M. Ritz, University of California, Santa Cruz

Panel on Radio Millimeter and Submillimeter Observations from the Ground
Jeremiah K. Darling, University of Colorado, Boulder
Neal Evans, University of Texas, Austin
Christopher F. McKee, NAS, University of California, Berkeley
Anna Scaife, University of Manchester

State of the Profession and Societal Impacts Appendix
Marcel Agüeros, Columbia University
Rosanna Alegado, University of Hawaii, Manoa
Neta A. Bahcall, NAS, Princeton University
Edmund Bertschinger, Massachusetts Institute of Technology
Joan Centrella, West Virginia University
Alison Coil, University of California, San Diego
Kate Follette, Amherst College
Felicia Jones, KBR, Inc.
Philip (Bo) W. Hammer, Institute for Mathematical and Statistical Innovation
Roberta Marinelli, Oregon State University

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report nor did they see the final draft before its release. The review of this report was overseen by Charles F. Kennel, NAS, University of California, San Diego, and Rosalba Perna, State University of New York at Stony Brook. They were responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

Contents

Summary

We live in an extraordinary period of discovery in astronomy and astrophysics. Six Nobel Prizes have been awarded over the past decade alone for discoveries based on astronomical data (dark energy, gravitational waves, neutrino oscillations, the discovery of exoplanets, cosmology, supermassive black holes). Many of the ambitious scientific visions of the 2010 *New Worlds New Horizons*[1] (NWNH) decadal survey are being fulfilled, but momentum has only grown. We stand on the threshold of new endeavors that will transform not only our understanding of the universe and the processes and physical paradigms that govern it, but also humanity's place in it.

This report of the Committee for a Decadal Survey on Astronomy and Astrophysics 2020 (Astro2020) proposes a broad, integrated plan for space- and ground-based astronomy and astrophysics for the decade 2023-2032.[2] It also lays the foundations for further advances in the following decade. This is the seventh in a sequence of decadal survey studies in this field from the National Academies of Sciences, Engineering, and Medicine. This survey examines the program of record, providing advice on the major projects from prior surveys that are yet to be completed. It also lays out priorities for future investments driven by scientific opportunities. The recommendations in this report advance foundational activities that support the people who drive innovation and discovery, and that promote the technologies and tools needed to carry out the science. The report also recommends sustaining activities on a broad range of cost and timescales, as well as activities that enable future visionary projects by maturing them scientifically and technically. Finally, the recommendations set in motion the construction of frontier facilities that will change the view and understanding of the cosmos. The survey is bounded by plausible budget scenarios based on briefings from the sponsoring agencies—the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Department of Energy (DOE). Within these bounds, the survey aims high, reflecting this time of great scientific promise and progress, with opportunities to pursue some of the most compelling scientific quests of our times.

THE SCIENTIFIC OPPORTUNITIES

The survey's scientific vision is framed around three broad themes that embrace some of the most exciting new discoveries and progress since the start of the millennium, and that promise to address some of the most fundamental and profound questions in our exploration of the cosmos. The first theme, *Worlds and Suns in Context* builds on revolutionary advances in our observations of exoplanets and stars and aims to understand their formation, evolution, and interconnected nature, and to characterize other solar systems, including potentially habitable analogs to our own. *New Messengers and New Physics* will exploit the new observational tools of gravitational waves and particles, along with temporal monitoring of the sky across the electromagnetic spectrum and wide-area surveys from the ultraviolet and visible to microwave and radio to probe some of the most energetic processes in the universe and also address the nature of dark matter, dark energy, and cosmological inflation. Research in the third theme, *Cosmic*

[1] National Research Council, 2010, *New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/12951.

[2] The statement of task specified a date range of 2022-2032. This has been adjusted to more accurately reflect the range that the Survey will affect.

Ecosystems, will link observations and modeling of the stars, galaxies, and the gas and energetic processes that couple their formation, evolution, and destinies.

Within each of these broad and rich scientific themes, three priority areas motivate recommended investments over the coming decade. "Pathways to Habitable Worlds" is a step-by-step program to identify and characterize Earth-like extrasolar planets, with the ultimate goal of obtaining imaging and spectroscopy of potentially habitable worlds. "New Windows on the Dynamic Universe" is aimed at combining time-resolved multi-wavelength electromagnetic observations from space and the ground with non-electromagnetic signals to probe the nature of black holes, neutron stars, the explosive events and mergers that give rise to them, and to use signatures imprinted by gravitational waves to understand what happened in the earliest moments in the birth of the universe. "Unveiling the Drivers of Galaxy Growth" is aimed at revolutionizing our understanding of the origins and evolution of galaxies, from the nature of the tenuous cosmic webs of gas that feed them, to the nature of how this gas condenses and drives the formation of stars.

THE RECOMMENDED PROGRAM

Major leaps in observational capabilities will be realized in the coming decade when new large telescopes and missions commence science operations (Table 7.1). Recommended by previous surveys, with some undertaken with international partners, these projects and programs are an essential base upon which the survey's scientific vision is built. It is essential that these initiatives be completed, and the scientific programs be supported at levels that ensure full exploitation of their potential by the U.S. scientific community.

Going forward, this survey lays out a strategy for federal investments aimed at paving a pathway from the foundations of the profession to the bold scientific frontiers.

Large Programs that Forge the Frontiers

These scientific visions—Pathways to Habitable Worlds, New Windows on the Dynamic Universe, and Unveiling the Drivers of Galaxy Growth— require the major recommended investments in large projects to begin design and construction in the coming 10 years (Tables S.5 and S.6; Figure S.1).[3] In space, achieving the community's most ambitious and visionary ideas in a sustainable way, and realizing the broad capabilities demanded by the richness of the science, requires a re-imagining of the ways in which large missions are planned, developed, and implemented. The **Great Observatories Mission and Technology Maturation Program** (Table S.5) would provide significant early investments in the co-maturation of mission concepts and technologies, with appropriate decadal survey input on scope, and with checks and course corrections along the way. Inspired by the vision of searching for signatures of life on planets outside of the solar system, and by the transformative capability such a telescope would have for a wide range of astrophysics, the survey recommends that the first mission to enter this program is a **large (~6 m aperture) infrared/optical/ultraviolet (IR/O/UV) space telescope.** The scientific goals of this mission, when achieved, have the potential to profoundly change the way that human beings view our place in the universe. With sufficient ambition, we are poised scientifically and technically to make this transformational step. This endeavor represents a quest that is on the technical forefront, is of an ambitious scale that only NASA can undertake, and it is one where the United States is uniquely situated to lead the world. If maturation proceeds as expected, the survey recommends that formulation and implementation begin by the end of the 2020 decade. To prepare for future large, strategic missions, 5 years after beginning the maturation program for the IR/O/UV mission, the survey

[3] For space, large projects are defined as those with costs exceeding $1.5 billion. For ground-based initiatives, large projects are defined as those exceeding $130 million for the total program investment.

recommends commencing mission and technology maturation of both a far-IR and an X-ray large strategic mission, both scoped to have implementation costs in the $3 billion to $5 billion range.

Because of the powerful potential that large (20–40 m) telescopes with diffraction-limited adaptive optics have for astronomy, and because of the readiness of the projects, the survey's priority for a frontier ground-based observatory is a significant U.S. investment in the Giant Magellan Telescope (GMT) and Thirty Meter Telescope (TMT) projects, ideally as components of a coordinated **U.S. Extremely Large Telescope Program (ELT) program**. These observatories will create enormous opportunities for scientific progress over the coming decades and well beyond, and they will address nearly every important science question across all three priority science areas. After this, given technical and scientific progress over the last decades, ground-based cosmic microwave background (CMB) studies are poised in the next decade to make a major step forward, and the **CMB Stage 4 (CMB-S4) observatory** (with support from NSF and DOE) will have broad impact on cosmology and astrophysics. It is also essential to astronomy that the Karl Jansky Very Large Array (JVLA) and Very Long Baseline Array (VLBA), which have been the world-leading radio observatories, be replaced by an observatory that can achieve roughly an order of magnitude improvement in sensitivity compared to those facilities. **The Next Generation Very Large Array (ngVLA)** will achieve this, with a phased approach where design, prototyping, and cost studies are completed and reviewed in advance of commencing construction. Finally, neutrino observations are important to understanding some of the most energetic processes in the universe, and the Ice Cube-Generation 2 (IceCube-Gen2) observatory will make advances in important astrophysics questions, although it is beyond the charge of this survey to recommend it.[4]

Programs that Sustain and Balance the Science

Turning to medium-scale missions and projects, the scientific richness of a broader set of themes—exploring *New Messengers and New Physic*s, understanding *Cosmic Ecosystems*, and placing *Worlds and Suns in Context*—as well as the need to capitalize on major existing investments and those coming online in the next decades drive the essential sustaining projects (Tables S.5 and S.6). In space, the highest-priority sustaining activity is a **space-based time-domain and multi-messenger program** of small and medium-scale missions. In addition, the survey recommends a new line of probe missions to be competed in broad areas identified as important to accomplish the survey's scientific goals. For the coming decade, a far-IR mission, or an X-ray mission designed to complement the European Space Agency (ESA's) Athena mission, would provide powerful capabilities not possible at the Explorer scale. With science objectives that are more focused compared to a large strategic mission, and a cost cap of $1.5 billion, a cadence of one probe mission per decade is realistic. The selection of a probe mission in either area would not replace the need for a future large, strategic mission. For ground-based projects, the highest-priority sustaining activity is **a significant augmentation and expansion of mid-scale programs**, including the addition of strategic calls to support key survey priorities. The survey also strongly endorses investments in **technology development for advanced gravitational wave interferometers**, both to upgrade NSF's Laser Interferometer Gravitational-Wave Observatory (LIGO), and to prepare for the next large facility.[5]

---

[4] IceCube is supported and managed by the NSF Division of Physics, rather than the Division of Astronomical Sciences.

[5] Technology development for gravitational wave detection is funded out of the NSF Division of Physics. The survey strongly endorses the importance of the science to astronomy and astrophysics.

Foundational Activities

A successful decadal survey strategy requires serious attention to the smaller but vital investments in the foundations of the research. The people who make up the profession are the most fundamental component of the research enterprise, without whom the ambitious facilities, instruments, and experiments, as well as the promised transformative discoveries, would lie unfulfilled. Recognizing that diversity is a driver of innovation, and that the astronomy and astrophysics enterprise can be at its most innovative only when it maximizes and fully utilizes the broadest range of human talent, the survey forwards several crucial programs (Table S.1) to support early-career entrants, with a strong emphasis on broadening access, removing barriers to participation, and creating an environment that eschews harassment and discrimination of all kinds. The future of the field also requires that greater attention be paid to issues of sustainability and accountability, and several recommendations address these issues. Among the recommendations regarding the state of the profession, the most urgent need is maintenance of accurate data on funding outcomes, because it is sufficiently critical to the other recommendations. (Table S.1).

Science cannot progress without the essential support to individual investigators who take the data and transform them into scientific understanding and discovery. Accordingly, augmenting the NSF Astronomy and Astrophysics Grants program is the highest priority among the foundational recommendations. Science also cannot progress without the necessary tools, such as archives, data pipelines, laboratory work, and theoretical tools that provide the essential, cross-cutting foundations. The computational revolution continues to transform the conduct and culture of astronomy through the growing roles of large surveys and shared public data sets, big-team research, applications of machine learning, and numerical simulations, among others, and research investments will need to evolve to adapt to this changing landscape. Several critical areas require a healthier balance in order to optimize the scientific returns on past and future major investments (Table S.2).

The currently operating facilities on the ground and in space, along with the scientists who use them, are the primary engines of scientific discovery and progress in astronomy and astrophysics. In this regard, it is essential to adequately support the costs of operating facilities in space and on the ground, review them regularly during their productive lives, and for ground-based observatories, maintain them as premier facilities with modern, state-of-the-art instrumentation. Table S.3 summarizes this report’s recommendations relative to the agencies’ operational portfolios.

A balanced portfolio that includes a healthy investment in small- and medium-scale projects that are competed, draws from the ingenuity and breadth of the community, and enables science on a broad range of costs and timescales is essential for sustaining a vibrant astronomy and astrophysics program. These activities sustain scientific progress, amplify and enhance return from operating missions and observatories, and respond nimbly to new discovery. The survey recognizes the foundational need for supporting basic technology development and the crucial role small- and medium-scale projects play in broadening science and as a means of developing the next generation of technologists and instrumentalists. Table S.4 summarizes recommendations aimed at strengthening these.

Enabling Future Visions

The community’s most ambitious and visionary ideas now require timelines that are pan-decadal and even multi-generational. This is particularly true for NASA’s large strategic missions and NSF’s premier observatories that are driven by transformative scientific visions but are technically challenging. They also represent large investments of resources. Optimizing the cadence of major facilities and developing them in a sustainable way that ensures the appropriate level of maturity prior to a decadal or agency commitment and tighter control on ultimate project costs requires new, enabling programs and approaches. The Great Observatories Mission and Technology Maturation Program would provide a new approach for developing large space strategic missions. In addition, for all large projects, the survey

provides decision rules and recommends reviews, where required, to ensure technical, scientific, and cost-readiness prior to commitment of major resources. The survey also identifies a few future projects that are targets for significant investment in maturation for consideration by future decadal surveys, as summarized in Tables S.5 and S.6, column 2.

A very large fraction of the astronomical community contributed to this survey through the almost 900 excellent science, activity, program, and state of the profession white papers and through active engagement in town hall meetings. The program laid out in this report represents a collective vision for the future and will require the engagement of a broad community to advance.

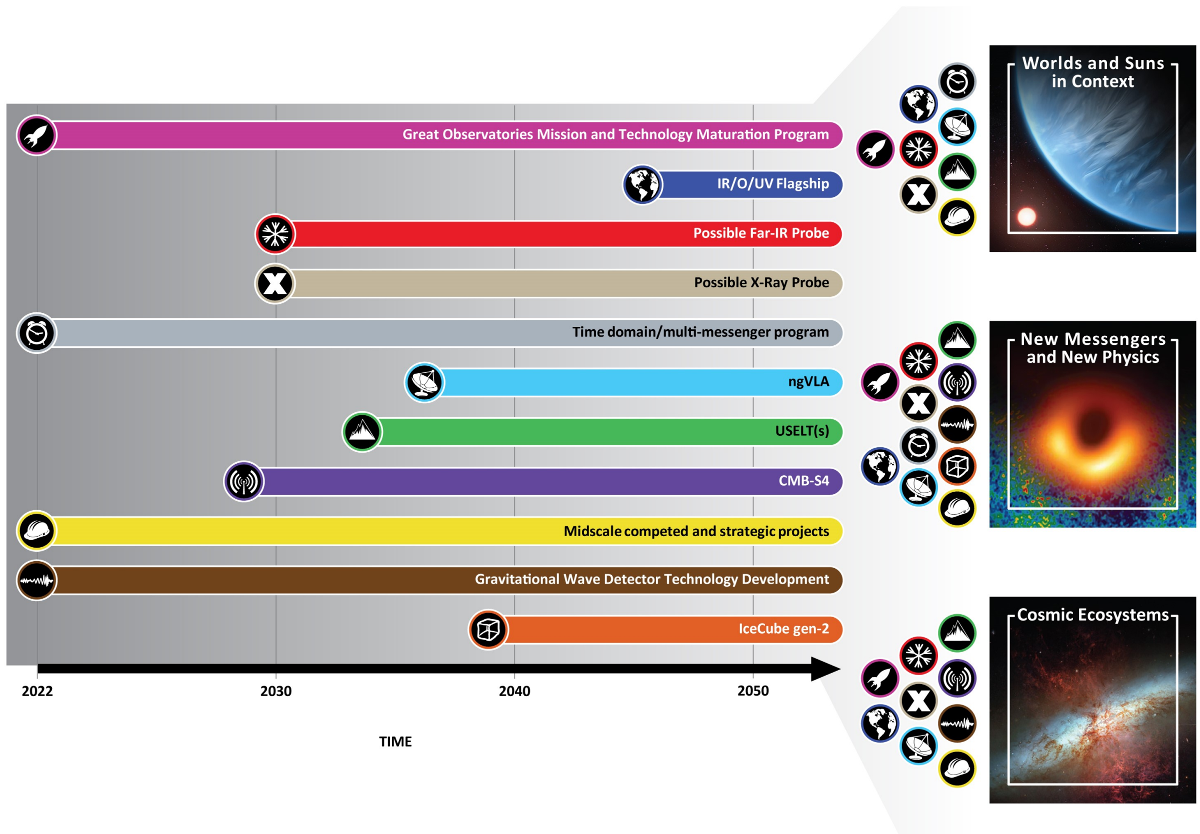


FIGURE S.1 Timeline for the recommended medium and large programs and projects. The starting point of each, indicated by the logos, shows the projected start of science operations for missions and observatories, or the start date of the program. The boxes on the right show the survey's three broad science themes, and the placement of the logos to the left of the boxes indicate which activities address the indicated theme.

Tables S.1 to S.6 below summarize the survey's recommended program, divided into tables that follow the chapter structure of the report. These tables are only intended to provide a capsule summary of the recommendations. The survey's report provides detailed guidance on the implementation of major programs and emphasizes the range of scales and capabilities necessary for a healthy, balanced, and visionary program. The ordering of projects in the tables below does not indicate priority ranking. The body of the report provides guidance on which programs or projects are the most urgent and have highest

priority within their programmatic category, emphasizing that even within a given cost scale, a balance of programmatic function is required.

TABLE S.1 Foundations of the Profession

Recommendation Topic	**Agency**	**Per Year Budget *Increases Relative to FY2019 Agency Budget Allocations (FY2020$)***	**Cross-Reference in Chapter 3**
Faculty diversity, and early-career faculty awards	NSF NASA DOE	(augmentation of) **$2.5M:** $1M NSF; $1M NASA; $0.5M DOE	p. 3-14
Workforce development/diversity, bridge programs and minority serving institutions partnerships	NSF NASA	(augmentation of) **$4.5M:** $1.5M NSF; $3M NASA	p. 3-22
Undergraduate and graduate "traineeship" funding	NSF NASA DOE	(augmentation of) **$3M:** $1M NSF; $1M NASA; $1M DOE	p. 3-23
Independent postdoc fellowships	NSF NASA	(augmentation of) **$1M:** $0.5M NSF; $0.5M NASA	p. 3-23
Treat discrimination and harassment as professional misconduct	NSF NASA DOE	N/A	p. 3-27
Collecting, evaluating, and reporting demographic data and indicators pertaining to equitable outcomes	NSF NASA	(augmentation of) **$1M** Split NSF/NASA	p. 3-29
Include diversity in evaluation of funding awards	NSF NASA DOE	N/A	p. 3-30
Establish community astronomy model for observatory sites	Community	N/A	p. 3-35
Mitigation of radio-frequency and optical interference from sources including satellite constellations	NSF NASA	TBD after evaluation	pp. 3-38 and 3-40
Climate change mitigation actions	Community	N/A	p. 3-42

TABLE S.2 The Research Foundation

Recommendation Topic	**Agency**	**Per Year Budget *Increases Relative to FY2019 Agency Budget Allocations (FY2020$)***	**Cross-Reference Page Number, Ch. 4**
Compile and regularly report data on proposal submissions and success rates	NSF NASA DOE	N/A	4-3
Augmentation to NSF Astronomy and Astrophysics Grants program	NSF	(augmentation) ramps up to additional $16.5M/yr (FY20) by 2028	4-8
Augmentation and restoration of annual proposal calls for Astrophysics Theory Program	NASA	(augmentation) ramps up to additional $2.5M/yr (FY20) by 2028	4-10
Support for large key projects on MREFC facilities	NSF	N/A	4-11
Improve coordination among U.S. data centers supported by NSF and NASA	NSF NASA	TBD depending on outcome of study	4-20
Data pipeline development, archiving for ground-based telescopes	NSF	TBD depending on plan	4-21
Augmentation and improved coordination of laboratory astrophysics funding	NSF NASA	(augmentation of) ~$2 M/yr, TBD after plan is developed	4-28

TABLE S.3 Sustaining the Operating Portfolio

Recommendation Topic	**Agency**	**Budget (FY2020$)**	**Cross-Reference Page Number, Ch. 5**
New MREFC facilities contingent on development of plan for supporting operations and maintenance costs	NSF	N/A	5-5
NSF to establish regular cadence of portfolio reviews of operating facilities	NSF	N/A	5-6
End SOFIA operations by 2023, consistent with current NASA plan	NASA	No impact if adopted	5-12

TABLE S.4 The Technological Foundations

Recommendation Topic	**Agency**	**Per Year Budget *Increases Relative to FY2019 Agency Budget Allocations (FY2020$)***	**Cross-Reference Page Number, Ch. 6**
Augmentation to NASA Astrophysics Research and Analysis program	NASA	(augmentation of) $4M/yr	6-4
Continue NASA Strategic Astrophysics Technology (SAT) program, expand eligibility to Explorer and Probe mission development	NASA	N/A	6-5
Augmentation to NSF Advanced Technologies and Instrumentation (ATI) program	NSF	(augmentation of) $8M/yr starting 2023, ramp up to $14M additional by 2028 (assumes current budget is $6M/yr)	6-6
Review NASA's balloon program for optimal balance	NASA	TBD depending on outcome of review	6-8-

TABLE S.5 New Medium and Large Initiatives: Space

Recommendation Topic	**Programmatic Function**	**Cost Appraisal (FY2020$)**	**Cross-Reference Page Number Ch. 7**
Great Observatories Mission and Technology Maturation Program for IR/O/UV (first half of decade), far-IR and X-ray (second half of decade) missions	Enabling future frontier projects	$1.2B this decade	7-11
Near-Infrared/Optical/Ultraviolet telescope with high-contrast imaging capability	Frontier project, to begin after maturation program	$11B (estimated)	7-17
Time Domain and Multi-messenger Follow-Up Program	Sustaining scientific balance and scale	TBD ($500-800M this decade est.)	7-19
Astrophysics Probe Mission Program	Sustaining scientific balance and scale	$1.5B cost cap	7-20

TABLE S.6 New Medium and Large Initiatives: Ground

Recommendation Topic	**Programmatic Function**	**Capital Cost** (FY2020$) (TRACE)	**Operations Cost** (FY2020$)	**Cross-Reference Page Number Ch.7**
Extremely Large Telescope Program (ELT)	Frontier Project	$1.6B (NSF share of $5.1B total project cost)	$32M/yr (NSF share of the $98M total)	7-24, 7-25
Stage 4 Cosmic Microwave Background Observatory (CMB-S4; joint NSF/DOE)	Frontier Project	$660M DOE+NSF; NSF share $273M	$17M/yr (NSF share of $40M/yr)	7-26
Next Generation Very Large Array (ngVLA)	Enabling development program, followed by construction if possible	$2.5B (NSF share of $3.2B project cost)	$98M/yr; NSF Share $73M/yr	7-28
Augmentation of Mid-scale Program: open and strategic calls	Sustaining	Ramps up to $50M/yr total for Mid-scale Innovations Program and Mid-scale Research Infrastructure	Operations in total program funding	7-29, 7-30
Technology development for gravitational wave LIGO upgrades and for future observatories	Enabling development program for future frontier GW observatories	N/A	N/A (not NSF AST)	7-31
IceCube-Gen2	Frontier project	N/A	N/A (not NSF AST)	7-32

1

Pathways to Discovery: From Foundations to Frontiers

We live in a time of extraordinary discovery and progress in astronomy and astrophysics. Since the dawn of the millennium, breakthroughs have come at an astounding rate, with highlights that include the first direct detection of gravitational radiation from astronomical sources; the discovery of thousands of extrasolar planets, including potential Earth-like analogs and the first characterizations of the physical properties and atmospheres for gaseous giant planets; mapping of the nascent disks of other solar systems as they are forming; a unified paradigm for the formation and evolution of galaxies, including deep insights gained from the fossil record of the Milky Way Galaxy; precision measurements of the supermassive black hole in the Milky Way's center; the first direct image of the shadow of a supermassive black hole; and precision measurements of the dark contents of the universe itself. Six Nobel Prizes for discoveries made using astronomical data have been awarded over the past decade alone (dark energy, gravitational waves, neutrino oscillations, the discovery of exoplanets, cosmology, supermassive black holes). Many ambitious scientific visions have been fulfilled in the past 10 years, but, if anything, momentum has only grown.

Every decade, the agencies that provide primary federal funding for astronomy and astrophysics—the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Department of Energy (DOE) Office of Science—request a decadal survey to assess the status of, and opportunities for the nation's efforts to forward our understanding of the cosmos. The National Academies of Sciences, Engineering, and Medicine responds by convening a body of experts with diverse interests and expertise to undertake this task, with a resulting report that advises the agencies about how to best deploy resources to advance knowledge in these areas. This survey's key objective is to map the national and international scientific landscape and to chart a path for investment, identifying programs with transformational scientific potential and new observational capabilities. Also central to the survey's charge is to assess the health of the profession and the balance of investments in the people and scientific infrastructure crucial to advancing the understanding of the cosmos. This report lays out a strategy for federal investments aimed at paving a pathway from the foundations of the profession to the bold scientific frontiers.

This chapter provides an integrated view of the strategy, analysis, and advice contained in Chapters 2-7. It is not a comprehensive summary of the report, but rather describes the recommended program in the broader context and framework in which this decadal survey was conducted, articulating the approach for building a scientifically broad, balanced, sustainable program that seizes the opportunities before us.

1.1 THE SCIENTIFIC OPPORTUNITIES

We stand on the threshold of new endeavors that will transform not only our understanding of the universe and the processes and physical paradigms that govern it, but also humanity's place in it. The tremendous richness of 21st century astrophysics is evident in the 573 science white papers authored by more than 4,500 individuals that lay out a wide array of questions we are now poised to answer. Six

expert science panels formulated these into key science questions and discovery areas ripe for rapid progress in the coming decade.

Three broad themes, described in Chapter 2, encompass these opportunities—*Worlds and Suns in Context*, *New Messengers and New Physics*, and *Cosmic Ecosystems.* The diversity of the science and observational techniques used to advance the associated goals is striking. Because of the balanced and varied programs put forward by prior decadal surveys, small telescopes, inventive experiments, and competed missions operating across the spectrum have harnessed the creativity and technical ingenuity of the community, resulting in an intensely dynamic and rapidly evolving enterprise. As a result, many of the questions at the forefront of the survey's themes could not have been framed even a decade ago. The richness of these three themes demands that a broad and varied suite of capabilities be sustained over the full electromagnetic spectrum and in the new windows of gravitational waves and high-energy neutrinos. Within each overarching theme, with its multiple science objectives, the survey identifies a priority science area that captures the most transformative and far-reaching goal, where, given new, ambitious facilities, we are poised to take giant strides forward.

1.1.1 Worlds and Suns in Context

The science theme of *Worlds and Suns in Context* captures the quest to understand the interconnected systems of stars and the worlds orbiting them, tracing them from the nascent disks of dust and gas from which they form, through the formation and evolution of the vast array of extrasolar planetary systems so wildly different than the one in which Earth resides. This is an area where advances over the past decade have been stunning, and progress in the next decade will be similarly rapid. By 2020, just 25 years after the discovery of the first exoplanet, the inventory of known exoplanets had exceeded 4,000, with more being identified nearly every week, thanks to ground-based radial velocity measurements and surveys of systems where the exoplanet partially eclipses its star (transit surveys), as well as dedicated space missions. The Kepler Discovery-class mission,[1] launched in 2009, revolutionized exoplanet studies by monitoring more than 150,000 stars to detect thousands of transiting planets, enabling astronomers to explore the structure and vast diversity of planetary systems for the first time. Combining Kepler's data with ground-based radial velocity measurements is providing essential information on exoplanet masses and densities. The Transiting Exoplanet Survey Satellite (TESS) Explorer-class mission,[2] launched in 2018, is surveying the entire sky to find nearby exoplanets, thereby providing the best sample for detailed follow-up studies using current and future ground- and space-based facilities. These same missions, along with the European Space Agency (ESA) Gaia astrometric and photometric observatory, launched in 2013, and large ground-based spectroscopic surveys have also enabled great leaps in the understanding of the physics of stars, the stellar populations of stars of the Milky Way, and the Milky Way's formation history.

The astronomical community and the public alike have been galvanized by the extraordinary progress in detecting and studying exoplanets. The 2018 National Academies report *Exoplanet Science Strategy*[3] captures this progress in rich detail. For the coming decade, key goals include applying spectroscopic and photometric observations to characterize exoplanet surfaces and atmospheres, and fully characterizing not only individual planets but also the properties of entire extrasolar planetary systems. The past decade has revealed how diverse and often different these are from our own solar system. But far more is needed to reliably assess the relative numbers of different system architectures. The upcoming Nancy Grace Roman Space Telescope, with launch expected in 2026, will conduct a microlensing survey

---

[1] The Discovery Program is a series of small to medium-sized competed solar system exploration missions funded by NASA Planetary Science Division.

[2] The Astrophysics Explorer Program is a series of small to medium-sized competed missions.

[3] National Academies of Sciences, Engineering, and Medicine, 2018, *Exoplanet Science Strategy*, Washington, DC: The National Academies Press, https://doi.org/10.17226/25187.

of the Milky Way's galactic bulge, filling out the census by finding exoplanets in the outer reaches of planetary systems that are inaccessible by other detection techniques. Ground-based 6–10 m optical and infrared telescopes with custom instrumentation will continue to broaden demographic samples and diagnose their properties. For the study of atmospheres of exoplanets in close-in orbits, as were found in abundance by Kepler and TESS, spectroscopy with the James Webb Space Telescope (JWST), to be launched by the end of 2021, will be transformational. Millimeter, radio, and infrared observations of the gas and dust disks of forming protoplanetary systems are providing complementary clues to the factors shaping the extent and architectures of solar systems, and this is an area of great discovery potential.

A rich agenda of discovery and scientific opportunity also lies ahead for stellar astrophysics. Over the coming decade, attention will focus on the most important unanswered questions, including understanding the effects of stellar multiplicity on the evolution of the stars in the system, the nature of stellar activity and activity cycles, and reconstructing the formation and assembly of the Milky Way as derived from its ancient stars. Precise distances, until recently only available for ~100,000 stars, are now available for hundreds of millions of stars, along with high-precision photometry, thanks to the ESA Gaia mission. With such a large sample, even rare types of stars and short-lived stellar evolutionary stages are well represented. At the same time, precision time-domain measurements of thousands of stars from Kepler, TESS, and the National Centre for Space Studies (CNES)/ESA Convection, Rotation and planetary Transits mission (CoRoT) have provided detailed asteroseismological measurements of their oscillations, which, like seismic measurements on Earth, unveil the internal structures and motions of material. Ground-based spectroscopy of the stars measured by the space missions will be crucial to obtain orbital velocities, chemical compositions, surface gravities, masses, rotation rates, and other fundamental properties. Spectroscopic survey telescopes in the 4-10 m class capable of observing thousands of stars simultaneously promise major advances. Finally, the Daniel K. Inouye Solar Telescope (DKIST) will revolutionize observations of the Sun's atmosphere.

Priority Area: Pathways to Habitable Worlds

Over the past two decades, thousands of extrasolar planets have been discovered, almost all of them extremely different from any world in our own solar system. This decadal survey's science theme of *Worlds and Suns in Context* encompasses the interlinked studies of stars, planetary systems, and the solar system. Within this broader science theme, the survey has identified the priority science area of Pathways to Habitable Worlds with the goal of trying to discover worlds that could resemble Earth and answer the fundamental question: "Are we alone?" Such planets will be found in the "habitable zone" of their parent stars—not too close and hot and not too distant and cold. Measurements indicate that around 30 percent of stars possess such a planet. The task for the next decades will be finding the easiest of such planets to characterize, and then studying them in detail, searching for signatures of life.

Life on Earth has profoundly altered the planet's atmosphere (Figure 1.1). Interpreting such "biosignatures" is not simple, but the interplay of atmospheric components such as water, oxygen, methane, and carbon dioxide can be modeled to search for evidence of life on other planets. Astronomers have already demonstrated the ability to use spectroscopy to study the atmospheres of large, hot worlds; with future facilities, the same techniques will measure the composition of small, habitable planets.

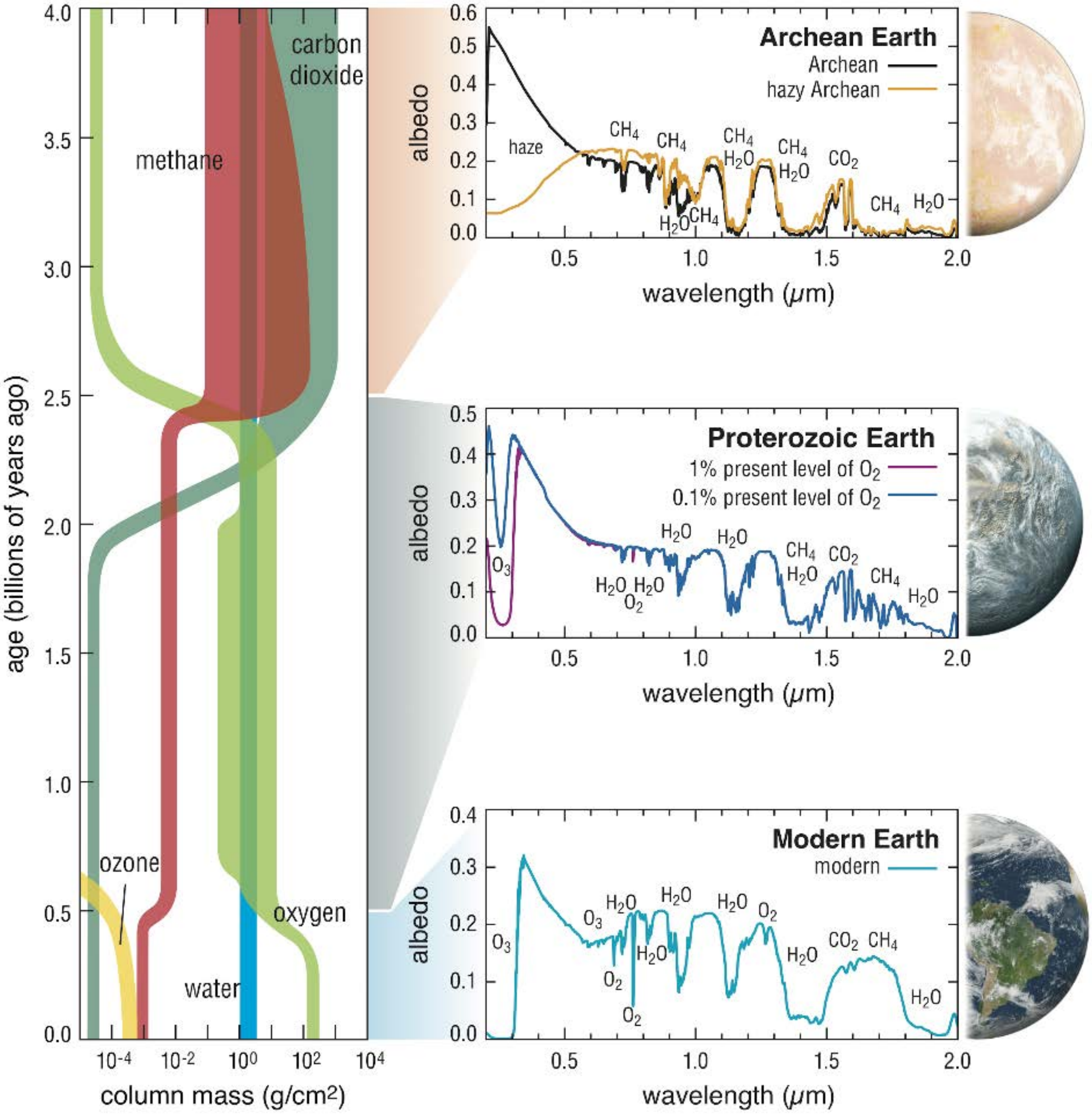


FIGURE 1.1 Evolution of the reflectivity spectrum of Earth. Simulated spectra of Earth before life had significantly altered its atmosphere (*top, Archean era 2.5 to 5 Gyr ago*), before the development of complex life (*middle, Proterozoic era from 0.54 to 2.5 Gyr ago*), and the modern oxygen-bearing Earth (*bottom*). SOURCE: LUVOIR Report; G. Arney, S. Domagal-Goldman, T. B. Griswold (NASA GSFC).

The pathway to searching for biosignatures on habitable worlds depends strongly on the properties of their parent stars. The most common stars in the Milky Way Galaxy are dim, red "M dwarfs." Their habitable zone will be very close to the star, making the systems accessible with the transit technique. JWST will observe a few of the very best target systems. To expand that sample will require the spectroscopic sensitivity of ground-based 25-40 m extremely large telescopes (ELTs).

However, M dwarf stars may not be the best harbor for life—they have massive super-flares and intense, potentially life-destroying energetic emissions. The planets around more placid Sun-like stars are essentially inaccessible to the transit technique and beyond the reach of ELTs, which must observe

through Earth's atmosphere. Only an ultra-stable, space-based telescope equipped to block the star's light and directly image the planet can reach this level of sensitivity. The larger the telescope, the larger the number of stars whose planets can be searched for signatures of life.

Properly interpreting these observations will also require a scientific context—understanding the formation and history of these planetary systems to see how life-enabling chemicals flow onto worlds, laboratory studies and simulations of planetary atmospheres, and deeper knowledge of the stars themselves—driving a large part of the overall *Worlds and Suns in Context* theme.

Key capabilities required on the pathway to habitable worlds include the following:

- Ground-based extremely large telescopes equipped with high-resolution spectroscopy, high-performance adaptive optics, and high-contrast imaging;
- A large, stable, space-based infrared/optical/ultraviolet (IR/O/UV) telescope with high-contrast imaging capable of observing planets 10 billion times fainter than their star, and UV, visible, and near-IR exoplanet spectroscopic capabilities;
- A high spatial and spectral resolution X-ray space observatory to probe stellar activity across the entire range of stellar types, including host stars of potentially life-sustaining exoplanets; and
- Laboratory and theoretical studies of planetary formation, evolution, and atmospheres.

Life on Earth may be the result of a common process, or it may require such an unusual set of circumstances that we are the only living beings within our part of the galaxy, or even in the universe. Either answer is profound. If planets like Earth are rare, our own world becomes even more precious. If we do discover the signature of life in another planetary system, it will change our place in the universe in a way not seen since the days of Copernicus—placing Earth among a community and continuum of worlds. The coming decades will set humanity down a path to determine whether we are alone.

1.1.2 New Messengers and New Physics

Our understanding of the universe has been repeatedly transformed by looking at the sky in new ways, from exploiting the full range of electromagnetic phenomena, to making large-scale, high-cadence astronomical movies, to exploring the universe in non-electromagnetic messengers. This has led to remarkable progress in astronomy over the past century, including the ever-growing impact of astronomy on basic physics. The *New Messengers and New Physics* theme captures the key scientific questions associated with a broad range of inquiries, from astronomical constraints on the nature of dark matter and dark energy, to the new astrophysics enabled by combined observations with particles, neutrinos, gravitational waves, and light.

The unknown physical natures of dark matter and dark energy, both discovered through astronomical measurements, remain outstanding grand challenges in both physics and astronomy, and great observational progress will be made in the coming decade. Addressing these profound mysteries were prime motivations for the Roman Space Telescope, with a field of view 100 times that of the Hubble Space Telescope (HST); the NSF/DOE Vera C. Rubin Observatory, a wide-field 8.4 m telescope devoted to a decade-long mapping of the entire southern sky; as well as ESA's Euclid mission, with a planned launch in 2022. These telescopes are all poised to address the nature of dark energy through large optical and infrared surveys aimed at measuring the distribution of galaxies on large scales, and by detecting distant supernovae. These measurements will also provide a lasting astronomical legacy, with data that can be mined to answer a variety of foundational astronomical questions. High-sensitivity and wide-angle mapping of the cosmic microwave background (CMB) has the potential to create virtual 3D tomographic maps of the matter distribution between the young universe—when there were free electrons that could readily scatter CMB photons—and Earth. These measurements can also be used to map the cosmic structure by mass (rather than by light, which is the structure traced by Roman, Rubin, and Euclid through

the starlight of galaxies). Comparing the two will reveal vital information about the structure itself, its evolution, and the evolution of differences between the distribution of light and mass.

In the past decade, a new, perplexing inconsistency between the expansion rates of the universe (Hubble constant) measured from nearby stellar distance ladders versus the CMB and other cosmological yardsticks has also emerged. The latter could be an observational issue, but it could also conceivably point to a missing element of physics in the current cosmological model. New measurements of the Hubble constant made by combining gravitational wave signals with associated redshift measurements will be an entirely independent way to resolve (or confirm) this tension.

The power of near-continuous monitoring of large regions of the sky in the X-ray, gamma-ray, optical, infrared, and radio bands has been dramatically demonstrated over the past two decades. Time-domain astronomy is now a mature field central to many astrophysical inquiries, from diagnosing the wide array of stellar explosions, to exoplanet detection, probing stellar structure, and measuring dramatic and unexplained changes in the appearance of active galactic nuclei—the regions closely surrounding supermassive black holes. New phenomena such as fast radio bursts, besides being events of mysterious origin, can provide a means of probing the tenuous gas in and in between galaxies. Progress in this subject will accelerate dramatically with the commissioning of the Rubin Observatory and later in the decade with the launch of the Roman Space Telescope. In particular, Rubin's unique time domain mapping of the southern sky is expected to detect roughly 10 million variable events per night, providing optical color information necessary for rapid characterization and unique scientific inquiries.

The ground-based Laser Interferometer Gravitational Wave Observatory's (LIGO) discovery in 2015 of gravitational waves from a pair of merging 30 solar mass black holes is certainly one of the watershed moments in physics and astronomy of the last decades. Future upgrades of LIGO, the European Virgo interferometer, and the Japanese Kagra, together with the launch of the Laser Interferometer Space Antenna (LISA) low-frequency gravitational wave observatory in the early 2030s have tremendous promise to answer fundamental questions in physics and astronomy and to open vast new discovery space. Upgrades of NSF's IceCube high-energy neutrino detector will enable these nearly massless subatomic particles to be associated with individual astrophysical objects, probing extreme environments where particles are accelerated to near-light speeds. The recent addition of the entirely new messengers—gravitational waves and high-energy neutrinos—to time domain astrophysics provides the motivation for the survey's priority science theme within *New Messengers and New Physics*.

Priority Area: New Windows on the Dynamic Universe

This report's science theme of *New Messengers and New Physics* captures the broad array of science made possible by observing the sky in new ways. Within this theme, the decadal survey has identified the priority science area of "New Windows on the Dynamic Universe"—the study of neutron stars, white dwarfs, collisions of black holes, and stellar explosions using the complementary perspectives provided by the wide range of messengers from light in all its forms from radio to gamma rays, gravitational waves, neutrinos, and high-energy particles. In parallel to remarkable advances in observations with multiple messengers from the LIGO/Virgo/Kagra gravitational wave and the IceCube high-energy neutrino observatories, the combination of large detectors, big data, and software advances for handling that data continues to transform the previously static view of the sky to one with nearly daily movies. Future upgrades of ground-based gravitational wave facilities, together with the launch of LISA make this a high priority for discovering new physics, and making astronomical measurements that will change paradigms.

Just like our everyday experience benefits from combining the information provided by sight, sound, taste, and smell, so too observations with these complementary messengers open new ways of doing astronomy and new ways of testing the laws of physics. This will reshape the understanding of topics as diverse as the origin of the carbon in bones and the metal in phones, the history of the expansion

of the universe since the Big Bang, the life and death of stars, and the physics of black hole event horizons.

New, coordinated advances in several areas are required to unlock the workings of the dynamic universe. These include the following:

- A suite of small and medium-scale ground and space-based observational facilities across the electromagnetic spectrum to discover and characterize the brightness and spectra of transient sources as they appear and fade away.
- Ground-based 20-40 m optical-infrared telescopes and an IR/O/UV space telescope significantly larger than HST to see the light coincident with colliding neutron stars detected in gravitational waves—most of these are sufficiently distant to be undetectable with current facilities. These same telescopes will diagnose in exquisite detail the elements produced in stellar explosions.
- A sensitive next-generation radio observatory more powerful than the Very Large Array (VLA) to detect the jets of relativistic gas produced by neutron stars and black holes, including those in mergers observed by ground and space-based gravitational wave facilities.
- Next-generation CMB telescopes to search for the polarization signatures of gravitational waves produced in the infant universe.
- Upgrades to improve the sensitivity of current ground-based gravitational wave detectors, and development of technologies to enable next-generation facilities.
- Improvements in the sensitivity and angular resolution of high energy neutrino observatories.
- Strong software and theoretical foundations to numerically interpret the gravitational wave signals from merging compact objects to extract new physics in the extremes of density and gravity, and ensure easy user access to the wealth of data on the dynamic universe and to model and interpret astronomical sources whose physical conditions cannot be replicated in laboratories on Earth.

1.1.3 Cosmic Ecosystems

The universe is characterized by an enormous range of physical scales and hierarchy in structure, from stars and planetary systems to galaxies and a cosmological web of complex filaments connecting them. A major advance in recent years has been the realization that the physical processes taking place on all scales are intimately interconnected, and that the universe and all its constituent systems are part of a constantly evolving ecosystem. The seeds of galaxies were planted during the first moments of the Big Bang, and modern numerical simulations trace the gravitational growth of cosmic structure from 300,000 years after the Big Bang to the structures and galaxies seen today. The galaxies are ecosystems of their own, with further condensation of matter to form stars and planets balanced by “feedback” from stellar winds, outflows, and supernovae that return mass and energy to the gaseous environment. The supermassive black holes that form and grow within nearly all massive galaxies also play a key role in this feedback process. Unraveling the nature of this connection is one of the key science goals of the decade.

The time is ripe for major breakthroughs. JWST will provide definitive observations of the earliest stages of galaxy formation and evolution, and the histories of star formation, chemical enrichment, and feedback processes over cosmic time. The combination of wide-area observations of distant galaxies by the Rubin Observatory, Roman, and Euclid will provide imaging and spectral energy information for millions of galaxies, complementing the in-depth observations from JWST and HST.

The upcoming observations with JWST, the Rubin Observatory, and Roman will be profound but will not on their own be able to address the central problem of understanding how galaxies grow. Probing the heart of the galactic feedback process requires detecting and measuring the tenuous gases at the boundaries of galaxies and their intergalactic surroundings, the circumgalactic medium (CGM), where the

accretion and recycling of gas and metals from feedback processes take place. This goal motivates the priority area within *Cosmic Ecosystems*.

Priority Area: Unveiling the Drivers of Galaxy Growth

Processes on a wide range of time and length scales shape the behavior of most astronomical objects, from the scales of planet formation in disks around young stars to galaxies and clusters of galaxies. The science theme of *Cosmic Ecosystems* captures the interconnectedness of these astronomical systems across cosmic scales. Within this theme, the decadal survey has identified the priority science area of Unveiling the Drivers of Galaxy Growth. The allure of galaxies—to scientists and the public alike—stems from their diversity and complexity. Their rich internal structures and tremendous variety make understanding the origin of galaxies one of the most continuously compelling stories in astronomy. The past decades have seen a growing understanding of the origin of this complexity: gas flows into galaxies, fueling new generations of stars and the buildup of central black holes, but these same stars and black holes send matter back out, potentially shutting down any chance for new material to stream in. These processes must have profound effects on galaxies, but astronomers have only a tenuous grasp on the full coupling between the larger galaxy environment that holds the gas transiting in and out of a galaxy, and the properties of the galaxy itself.

This profoundly multiscale problem requires connecting galaxies from their central black holes, in a region no larger than the solar system, to their outermost reaches millions of light years from the center. Technologically, these demanding requirements drive investments in reaching high resolution—to uncover the parsec-scale astrophysics powering feedback—and towards high sensitivity—to both detect the most tenuous and diffuse emission and to allow spectroscopy against faint background light sources with sufficient density to sample a dozen or more lines of sight in a single galaxy. Furthermore, the range of gas temperatures (from more than a million degrees kelvin down to temperatures approaching absolute zero) and redshifts naturally motivates a multiwavelength approach.

New observational capabilities across the electromagnetic spectrum along with computation and theory are needed to resolve the rich workings of galaxies on all scales. These include the following:

- Large 20–40 m ground-based O/IR telescopes to observe the transition-rich rest-frame UV, in both emission and absorption, for galaxies in the young universe. This will reveal the faint networks of gas that surround galaxies and the gas's chemistry, temperature, density, and motions.
- A next-generation VLA radio telescope will, for the same early epochs, map emission lines of molecular gas, tracing the cold gas associated with both the extended galactic environment and fueling AGN and star formation within the galaxy itself.
- A next-generation IR/O/UV large space telescope to trace much of the same physics as the ELTs but in the nearby, evolved universe, and in dramatic detail, revealing the full multiphase complexity of the local ecosystem.
- To complement these capabilities a capable far-IR and/or X-ray mission will further transform these views by peering into the dusty hearts of galaxies to reveal enshrouded accreting black holes, or tracing the hottest phases of gas driven outward by this same accretion, with the spatial and spectral resolution needed to isolate critical physical quantities in massive galaxies.
- Investments in theory and in the community of scientific experts exploring these data are essential for synthesizing a new scientific foundation for galaxy evolution from these observational advances.

1.2 BACKGROUND AND CONTEXT

The range and variety of compelling scientific opportunities illustrates the dynamic nature of modern astrophysics, with future directions propelled both by steady evolution and by dramatic revolution, powered by new discoveries, emerging capabilities, and an increasingly diverse set of ideas. The survey's recommended program is driven by the science, but it is also shaped by the global landscape, and the scientific, technical, and human context of the times. Multifaceted considerations led to the balance of science, the emphasis on sustainable investments in projects and people, and the wide-ranging activities on all scales that are prioritized through recommendations in this report.

The major scientific progress in astronomy over the past decade has been mirrored by a continued transformation in the national and international landscape in which this research is being conducted. Astronomy continues to become more global and interconnected, and many of the major space missions in recent decades (HST, JWST, Herschel Space Observatory, and Planck) have been carried out as partnerships between NASA, ESA, and/or the Japan Aerospace Exploration Agency (JAXA). With the XRISM and Athena X-ray observatories, Euclid, and LISA on the horizon, the survey's scientific goals are crucially dependent on such partnerships continuing and even strengthening going forward. On the ground, the Gemini and ALMA ground-based observatories are international collaborations with NSF participation. This trend is likely to continue; a majority of the large ground-based projects presented to this survey have, or plan to have, significant international partners. Data produced by other European-led observatories such as the ESA Gaia mission and the European Southern Observatory have also contributed to major advances by U.S. researchers, either individually or as members of international collaborations. This international context of current and planned facilities has been fully incorporated into the survey's science and strategy planning.

The imminent launch of JWST is a momentous occasion that will shape the course of astronomy and astrophysics in the coming decades. Arguably the most ambitious robotic science mission that NASA has ever undertaken, JWST will influence essentially every area of astronomy, from peering back in time to view nascent galaxies as they begin to form in the early universe, to exploring the atmospheres of exoplanets in exquisite detail. JWST, more than two decades in the making, reminds us of the transformational nature of the ambitious, large strategic missions that NASA is uniquely capable of undertaking.

While large strategic missions are transformative, 21st century astrophysics owes much of its richness to NASA's panchromatic suite of Great Observatories that spanned the spectrum from gamma rays to infrared, and which were accomplished with a wide range of scales, from what today is referred to as "Probe scale" up to the very ambitious HST and JWST missions. Diverse missions of all scales, national and international, designed to view the universe in a multiplicity of complementary ways are now essential to progress in modern astrophysics. The broad science laid out in this report requires a wide variety of space-based techniques and capabilities spanning not just the electromagnetic spectrum, but, with the launch of ESA's LISA mission, in which the United States is a significant partner, the gravitational wave spectrum as well. While, as noted above, sustaining broad observational capabilities is crucially dependent on international partnerships and missions, essential capabilities, such as very high-contrast imaging and spectroscopy in the IR/O/UV bands, far-IR imaging and spectroscopy, and high-resolution X-ray imaging and spectroscopy, are not planned in ESA's Voyage 2050 program,[4] or by other international agencies. Because of the significant U.S. leadership in the development of the needed technologies and capabilities, and the high priority these have for this survey, it is essential for NASA to lead their development. However, without a major change in the approach to developing strategic missions, combined with expanding the range of mission scales, it will take many decades to realize the necessary range of observational capabilities.

---

[4] See https://www.esa.int/Science_Exploration/Space_Science/Voyage_2050_sets_sail_ESA_chooses_future_science_mission_themes.

On the ground, the astronomical community eagerly awaits the commissioning of the Rubin Observatory, which will be devoted to a decade-long mapping of the entire southern sky in multiple colors and with multiple time-domain cadences. By harnessing the power of the digital revolution, and building on past large surveys, large public data sets, data science, and computational astrophysics, Rubin will leave a legacy of data that will be mined far into the future by a diversity of astronomers. The challenge going forward is to ensure that the vast range of variable and transient phenomena that Rubin will uncover can be quickly discovered and studied by facilities spanning the wavelength spectrum.

Concerning new ground-based activities, NSF and DOE strongly urged the survey to be ambitious and challenged it to consider bold, transformative initiatives. At the same time, NSF Division of Astronomical Sciences (NSF AST) is faced with an historic underinvestment in smaller scale, foundational activities such as the general investigator grants that ensure high scientific return from projects of all scales. Together with the lack of a sustainable model for operating new facilities, the agency faces structural issues it must address to capitalize on the opportunities. Nationally, attention is growing on the country's urgent need to build its infrastructure, technological base, and scientific foundations, and this movement aligns well with NSF AST's needs. Being a field that captures the imagination of the public, pushes technology, and is a gateway to science, technology, engineering, and mathematics (STEM) education, astronomy is in a good position to argue for addressing these foundational issues through increased basic investments.

The activities and deliberations of this decadal survey took place in a time of tremendous national and international upheaval. The global COVID-19 pandemic has disrupted every aspect of life, from seemingly mundane issues of how to conduct the survey's business, to health, childcare, elder care, and education. The impacts have not been equally felt by women and men, and they also depend on socioeconomic status and race. The careers of many young people, including scientists, have been paused, and this will have a lasting impact on the profession. The pandemic also strongly underscored the important role of science, and scientific reasoning in combatting the epidemic, from the rapid development of mnRNA vaccines, to the factual, analytic presentation of the data necessary to design protective measures. The ultimate economic and social impacts of the pandemic remain unclear, adding to the uncertainty of the future landscape.

As a final, important backdrop, this survey was strongly influenced by the urgent need to advance diversity, equality, and inclusion in all aspects of society. This need came into sharp focus with the Black Lives Matter movement, sexual harassment and the inequalities highlighted by the #MeToo movement, the inequitable impacts of COVID-19, and the shocking increase in hate crimes against Asian Americans. These harsh realities have invigorated the nation into a renewed conviction to tackle systemic issues of race, gender bias, and privilege at a local and global scale. There is momentum to effect change, and the time is overdue to actively focus on these activities. Changing the defaults under which astronomy is practiced will only happen with energetic engagement and a diversity-, equity-, and inclusion-focused lens.

1.3 FRAMEWORK FOR THE SURVEY'S RECOMMENDATIONS

In the context described above, the decadal survey committee weighed many considerations in designing its recommended program (Figure 1.2). Primary among these is that the portfolio must be scientifically balanced, broad, and sustainable. Also, the program must be structured to draw from the widest range of human talent. The first consideration drives the need for a balance of investments among activities that lay the foundations of the science and the profession, and that advance a variety of projects on all scales. The second consideration requires that the profession and the agencies nurture, structure, and expand programs in such a way that they eliminate barriers, create inclusive environments, and actively encourage broad participation.

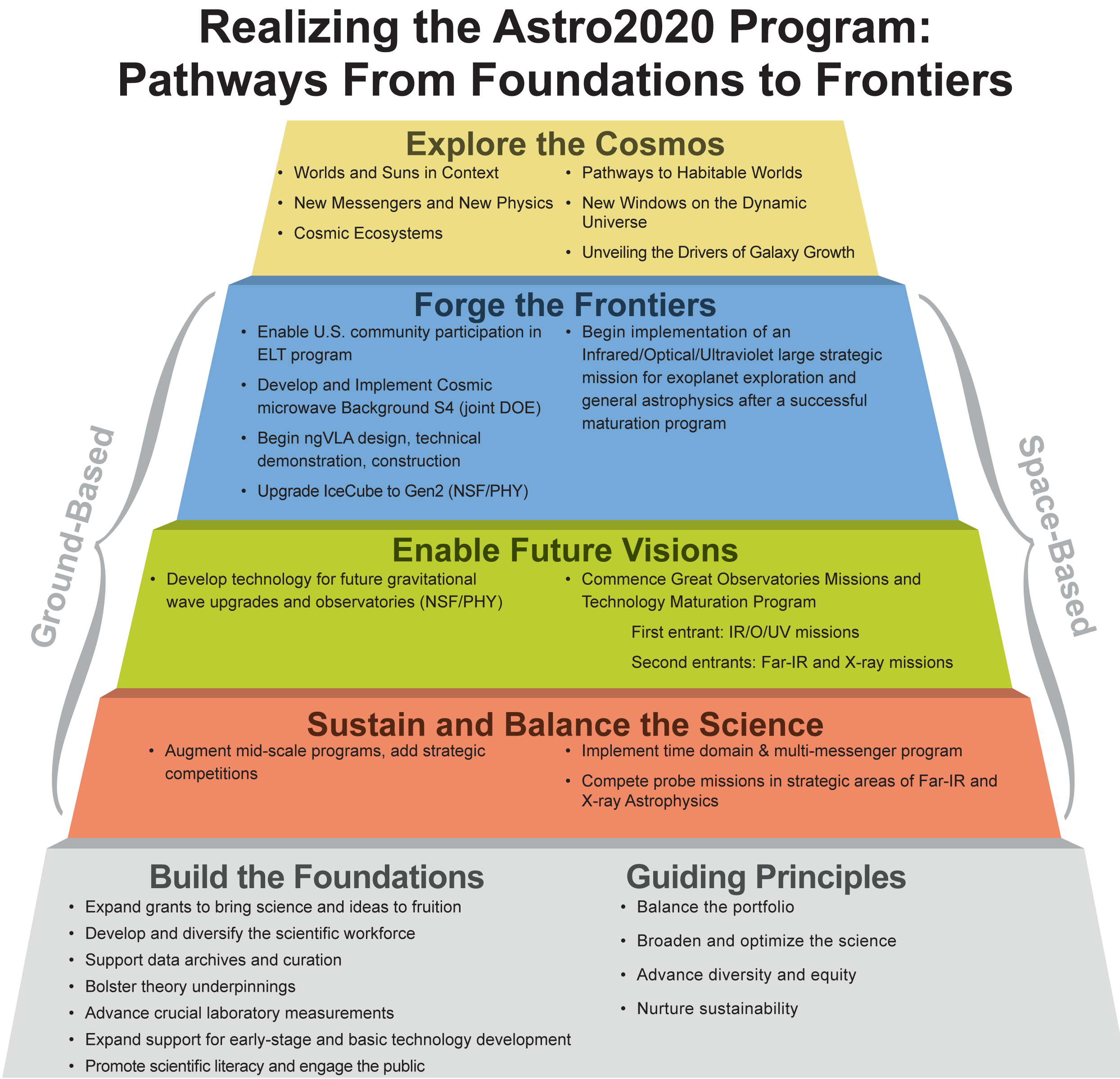


FIGURE 1.2 The recommended program includes elements that pave the way to transformative science by building a strong research and technology foundation, promoting programs on a range of scales that balance and sustain observational capabilities, enabling future large projects, and advancing new frontier observatories.

The survey's organization of projects and activities into categories is a departure from past practice. It emphasizes the function of the activity within the program rather than the cost, although there is a rough equivalence. Prior surveys have divided programs strictly by budgetary requirements (small, medium, and large) and have in general not prioritized projects in one cost category compared to those in another as a means of emphasizing the need for balance. The approach taken by Astro2020 is to adopt functional categories. Projects that build the foundations consist in large part of competed grants to individual investigators and programs that support modest scale activities, and sustaining projects consist

of competed mid-scale experiments and missions. Two categories capture the large, ambitious initiatives: programs that enable future visions and those that realize frontier facilities. This survey, like its predecessors, places strong emphasis on balance and the need for projects on a variety of scales and does not prioritize one category over another. In most categories, the survey identifies the highest-priority activity, and for others there is a natural time ordering based on scientific urgency and/or project readiness.

Another consideration is the budget uncertainty associated both with agency guidance, and with the landscape of federal funding discussed above. NASA and NSF strongly urged the survey to develop a program that is aspirational and inspirational, but that also conforms to budgetary norms. Given the uncertain landscape, the survey committee concluded that it is not possible to imagine and plan for the many possible contingencies. Rather, the recommended program forwards the frontiers of science through ambitious projects, and at the same time strongly advocates for balance. With this guidance in hand, the agencies have the flexibility to seize opportunities that arise on all scales, and the strong motivation to do so given the analysis of this report. For major projects that dominate budgetary requirements, the survey establishes decision rules and off ramps that guide agencies in the event of technical issues, or changes in the budgetary landscape. Interim advice from the mid-decadal survey, and from committees such as the Space Studies Board, the Board on Physics and Astronomy, and the Committee on Astronomy and Astrophysics, are an effective means for the agencies to request input on issues resulting from changing circumstances. These corrections would of course be strongly guided by, and be based on, the full analysis contained in this report. Additional prescriptions are unlikely to be helpful to the agencies given the many constraints, fiscal, political, and organizational, that they are faced with.

The greatest challenge faced by the survey committee in developing new recommendations for the nation's space astrophysics program is how to realize large strategic missions, yet at the same time achieve the wavelength balance, and the overlapping operational lifetimes that characterized NASA's Great Observatories, a model that so successfully propelled many, varied fields of astrophysics. While international partnerships are essential, they are not sufficient to accomplish the broad and aspirational science program laid out in Astro2020. Doing so will require a range of missions significantly larger than Explorers, yet with a mix of cost and implementation time scales spanning from a less than a decade to the multiple decades required to realize a mission of the ambition and complexity of JWST.

As evidenced by the four Large Mission Concept Studies prepared for this survey, the community's most ambitious and visionary ideas now require timelines that are pan-decadal, and even multi-generational (Chapter 7, Table 7.4). We are poised to tackle some questions that are so grand that the facilities and instruments needed to address them require vision and commitment beyond our individual horizons. But to do this sustainably, and to realize the broad capabilities demanded by the richness of the science requires a re-imagining of the ways in which large missions are developed and implemented. The ambitious strategic missions demand much more significant early investments in co-maturing mission concepts and technologies prior to adoption, with appropriate decadal input on scope, and with checks and course corrections along the way. In addition, adding a competed probe mission line that spans the large gap between Explorers and ambitious strategic missions, with science foci identified by decadal surveys will be a further move toward a capable, panchromatic mission fleet.

The greatest challenge for NSF going forward is its need to develop the appropriate programmatic balance of projects spanning the needed range of budgetary levels required to optimize the return on federal investments. Seizing the compelling scientific opportunities, and retaining U.S. competitiveness in astronomy requires capabilities that are uniquely provided by large, ambitious facilities. However, it also requires supporting operations of NSF's wide range of productive facilities, including upgrading instrumentation, ensuring a balance of project scales, and most importantly supporting the community of individual investigators to realize the scientific goals set out for the decade. The complex challenge associated with achieving this balance has been an impediment to the field for multiple decades, and it must be addressed if we are to reap the scientific rewards going forward.

For NASA, NSF, and DOE, overruns and delays in major projects have historically been a significant threat to improving and maintaining program balance. The survey addresses this in two ways.

First, the recommendations in this report emphasize more significant project and technology maturation prior to a commitment to, and commencement of, implementation. This will enable requisite budgets to be more firmly established when the project is adopted by the agency. In addition, for major projects, the decision rules are intended to guide the agencies in how to manage changing circumstances, technical or budgetary.

1.4 CRITERIA FOR ADVANCING NEW ACTIVITIES

The program of new activities in this report was conceived in the context of numerous exciting large strategic projects and missions, including international partnerships, that have yet to begin scientific operation (see Table 7.1 for a comprehensive list). This survey assumes that these compelling programs will be all be completed and sustained through their scientifically productive lifetimes. Ambitious and transformative large-scale efforts often take multiple decades to realize, and all of those scheduled for completion in the coming decade will provide essential capabilities upon which the Survey's scientific goals rely. Further, programs resulting from decadal recommendations, such as NASA's expanded Explorer program and NSF's Mid-Scale Innovations Program, play essential roles in sustaining scientific breadth and ensuring timely response to new opportunities. These continued and future capabilities are essential underpinnings upon which new recommendations are predicated.

For NSF, as noted above, the pressure imposed by operations costs of large NSF facilities on grants and other programs has been a systemic issue plaguing astronomy. By the middle of the decade, this will escalate to unsustainable levels unless changes are made to the way that large facilities are supported. The survey's recommendation is that new, large Major Research Equipment and Facilities Construction (MREFC) recommendations described below be predicated on NSF developing a sustainable plan for supporting the operations and maintenance costs of its astronomical facilities, while preserving an appropriate balance with funding essential scientific foundations and the remainder of the NSF AST portfolio.

1.5 RECOMMENDED PROGRAM OF NEW ACTIVITIES

The survey's recommendations for new programs and program augmentations are organized into steps that form the pathway from the foundations of the profession out to the scientific frontiers (Figure 1.2). The full text of the survey committee's analyses and recommendations is found in Chapters 2-7, while this chapter provides a broad overview. These recommendations advance transformative science in the coming decade and set the stage for enabling the bold visions in the future (Figure 1.3).

1.5.1 Guiding Principles

Major investments must advance a bold and broad scientific vision, while at the same time ensuring a balanced program that responds to scientific opportunity. Astronomy and astrophysics advances in a global context, and the survey recognized and responded to the need for synergy with, and complementarity to, activities worldwide. Especially for ground-based observatories, private institutions and philanthropic entities have been, and continue to be central to some of the most ambitious endeavors. The survey committee carefully considered how to best leverage these private-public partnerships in a way that achieves ambitious science and advances the aspirations of the entire community. There is also the challenging issue of balancing scientific ambition with feasibility and timeliness. All of these factors shaped the recommended programs, and their phasing.

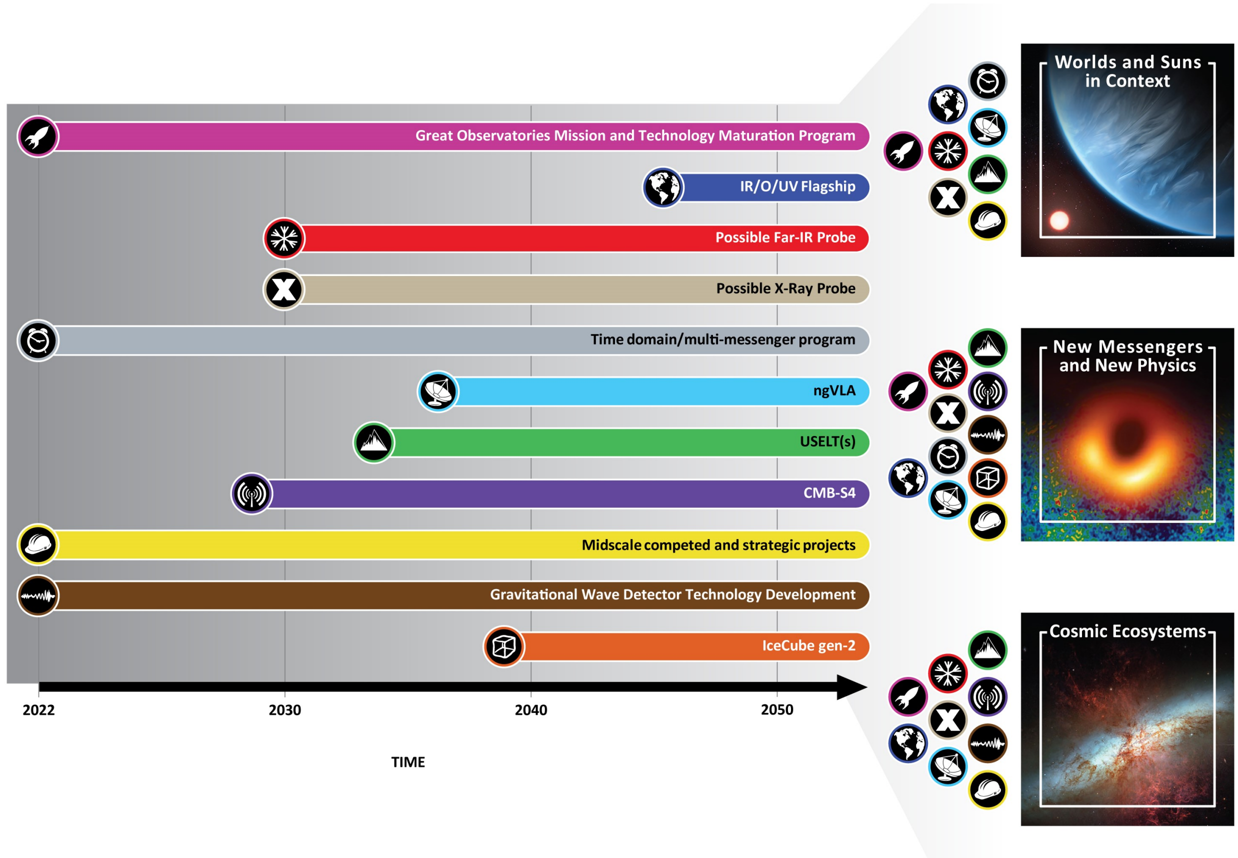


FIGURE 1.3 Timeline for the medium and large programs and projects recommended by this Astro2020 decadal survey. The starting point of each, indicated by the logos, shows the projected start of science operations for missions and observatories, or the start date of the program. The boxes on the right show the survey's three broad science themes, and the placement of the logos to the left of the boxes indicate which activities address the indicated theme. As evidenced in the figure, advancing each of the survey's broad science themes requires a range of facilities and programs.

The recommendations in this report are also guided by the precepts and principles of diversity, equity, benefit to the nation and the world, and sustainability. Diversity is a driver of innovation, and the astronomy and astrophysics enterprise can be at its most innovative only when it maximizes and fully utilizes the diversity of its human talent, ensures equitable access to opportunities, removes barriers to participation, and when it values diverse forms of expertise in its leadership. Equity demands that what is pursued with the nation's resources are done in a manner consistent with the principles of fairness and equal opportunity that are core to society's ideals. Anyone with the ability and the drive to contribute through astronomical discovery should have a fair chance to do so, and be free of fear, harassment, or discrimination. The benefits of astronomy and astrophysics extend beyond its fundamental discoveries. They provide lifelong learning opportunities and science literacy to the public and contribute to the development of the nation's broader, technically trained STEM workforce. In terms of sustainability and accountability, the substantial investments in people and the use of natural resources in astronomy require responsible stewardship, transparency, and accountability for outcomes. This is a core responsibility of the organizations, agencies, and stakeholders that benefit from the human labor and products of the field.

1.5.2 Foundational Activities

The pathway begins with strong support for foundational activities that build the people and the profession, (Chapter 3, The Foundations of the Profession), bolster the core activities necessary for a vibrant research enterprise (Chapter 4, The Research Foundations), and lay the technological foundations for the future (Chapter 6, The Technological Foundations and Small and Medium Programs that Balance the Science). The key new programs with these aims are described below.

- *Develop and diversify the scientific workforce.* The diversity of the astronomy and astrophysics profession remains an area where much improvement is needed. While there have been some notable improvements, especially with regards to the representation of women at the early career ranks of the profession, the overall demographics of the field remain very far from parity with the larger population. Addressing this will require action on many fronts: recommendations in this decadal survey report span the career stages from undergraduate to faculty and beyond, with targeted programs to improve diversity at each level; bridge critical transitions in the pipeline; and work to improve diversity of project teams, participants, and beneficiaries. The ugly realization of continued discrimination in the form of racism, bias, and harassment hampers progress towards building a fully diverse and inclusive workforce, and a recommendation of the report in this area suggests adoption of scientific integrity policies that address discrimination and harassment as forms of research or scientific misconduct. At the core of a diversity-, equity-, and inclusivity-focused approach is the need for data to evaluate equitable outcomes of proposal competitions; such data was sorely lacking in the preparation of this report, and a recommendation to collect, evaluate, and publicly report such data would enable future assessments.
- *Promote scientific literacy and engage the public*. By capturing the public's attention with discoveries, including the participation of citizen scientists in the research process, promoting science literacy, and realizing advanced technologies that can then find real-world applications, astronomy has a clear benefit to the nation. Astronomy education is effective as a broad gateway to STEM careers. Considering the rapidly increasing need for advanced computational skills in both the public and private sector for students to be competitive, embedding computational training in the undergraduate curriculum is even more important to integrate in the coming decade.
- *Promote sustainability and accountability*. The future of the field requires that greater attention be paid to issues of sustainability and accountability, whether it is in the context of the natural resources required for astronomy research activities at observing sites, or the current crisis of a large number of low Earth–orbiting satellites that will impact wide-field imaging at optical wavelengths and radio frequency observations. Adapting to the realities of climate change requires a decrease of the field's impact on carbon emissions. Recognizing the need for active, up-front, and sustained engagement with local and Indigenous communities, the survey committee recommends the implementation of a Community Astronomy model of engagement, similar to community-based approaches in other scientific disciplines. The goal for such an approach is to advance scientific research while also respecting, empowering, and benefitting local communities.
- *Expand the NSF grants program (highest-priority foundational recommendation).* Robust individual investigator grant funding is crucial to achieving the science goals of this decadal survey and to ensure more equitable access to resources. The NSF Astronomy and Astrophysics Grants (AAG) program is a cornerstone of the enabling foundation for research in astronomy and astrophysics in the United States, supporting research projects across nearly all subfields of the astrophysical sciences. This program is not currently at a healthy level, and the recommendation for an augmentation over 5 years is designed to restore success rates

to a healthy competitive environment. This is the highest-priority item amongst the many recommendations for building the foundation of the nation's research enterprise.

- *Bolster theory underpinnings.* Theoretical investigations, crucial as both a mechanism for driving new discoveries and a framework for interpreting essentially all signals received from space, are, like grants at NSF, lacking crucial funding at a level that can sustain the necessary projects. A recommendation to increase the amount of funding for NASA's Astrophysics Theory Program (ATP), and restore its cadence to an annual call, reflects increases to recover from past limited funding.
- *Maximize science from large programs on ground-based facilities.* Another survey recommendation in the foundations category urges NSF to establish a mechanism of research funding and production of high level data products for large principal-investigator programs on MREFC-scale astronomical facilities. This would accelerate scientific output and maximize the timeliness and community impact of large key projects.
- *Support data archives and curation.* Astronomy is evolving rapidly into a profession in which archiving of individual observations can produce scientific impacts that rival the original studies, and large-scale surveys are designed for science-ready archival manipulation from the beginning. As demonstrated by space missions and some ground observatories (e.g., ALMA, the European Southern Observatory [ESO]), readily-accessible archival data can substantially increase the scientific impact of facilities for a relatively modest incremental cost. The situation is less uniform for the large number of ground-based optical/infrared (OIR) facilities managed by universities and other institutions. A survey recommendation to NSF and stakeholders for enabling science-ready data across all general-purpose ground-based observatories is an attempt to ensure that all pipelined observations are archived for eventual public use.
- *Advance crucial laboratory measurements.* Laboratory astrophysics is a critical but often hidden and underappreciated cornerstone of the enabling research foundation. It has been chronically underfunded; concerns were raised in both the 2000 and 2010 decadal surveys, but the problem persists. Research in this area needs to be regarded as a high priority, and the existing approaches are not sufficiently advancing the field. A multi-step recommendation in this area urges the agencies to identify the needs for supporting laboratory data to interpret the results of new astronomical observatories, identify resources, and consider new approaches or programs for building the requisite databases. The recommendation also points out the need to include not only experts in laboratory astrophysics, but also users of the data to identify the highest priority applications.
- *Expand support for early-stage and basic technology development.* Analyses of the needs for basic technology funding to support future innovation, as well as to advance identified goals for, for example, high-contrast imaging, adaptive optics, highly multiplexed detectors, and technologies that will drive the next generation of instruments, observatories and missions, identifies increased investments in basic technology as a priority. Another important factor is that basic technology grants are too small to support infrastructure or significant involvement by industrial partners. To be able to fuel innovative future projects on all scales, it is important for the basic technology development portion of the Astrophysics Research and Analysis Program (APRA) to be significantly increased, and for cuts to NSF's Advanced Technologies and Instrumentation (ATI) program over the last decade be rapidly reversed, and additional funding be added to bring the program to the levels recommended by Astro2010.

1.5.3 Programs that Sustain and Balance the Science

Chapter 7 (The New Medium and Large Investments that Sustain Science and Forge Frontiers) lays out an ambitious roadmap for high-priority, space- and ground-based, large and medium-scale initiatives that are compelling and ready to begin implementation in the decade 2023–2033. This roadmap has at its core recommendations are aimed at capitalizing on the upcoming Roman, Rubin, Athena, and LISA observatories, and balancing scientific progress among the survey's priorities, thereby addressing the extraordinary richness of 21st century astrophysics.

Time Domain Astrophysics Program (Highest Priority Sustaining Activity for Space)

Exploring the cosmos in the multi-messenger and time domains is a key scientific priority for the coming decade, with new capabilities for discovery on the horizon with the Rubin Observatory, Roman, LIGO/Virgo and the Kamioka Gravitational Wave Detector (KAGRA), and IceCube. To advance this science, it is essential to maintain and expand space-based time-domain and follow up facilities in space. Many of the necessary observational capabilities can be realized on Explorer-scale platforms, or possibly somewhat larger. As the international landscape and health of NASA assets change, it will be important for NASA to seek regular advice over the coming decade on needed capabilities and to ensure their development. The open Explorer program calls have reached a healthy funding level, and as noted in Section 6.2.1.1.3, maintaining the current cadence of open calls is a condition for new initiatives. This time-domain program is therefore recommended as an augmentation to those levels, and would be executed through competed calls in broad, identified areas.

Astrophysics Probe Mission Program (Space)

The large gap in cost and capability between medium-class Explorer missions and the large strategic missions presented to the survey is a significant impediment to achieving the broad set of decadal scientific priorities. Institution of Probe-class line of missions with a cost cap of ~$1.5 billion per mission, a cadence of ~one per decade, and competed within selected priority areas identified by this and future decadal surveys, is a crucial addition to NASA's astrophysics portfolio. The two priorities for the first Probe-class mission competition are a far-IR probe or an X-ray probe to complement the Athena mission. Both areas represent important observational needs where advances in technology and focused objectives can yield transformative science on a moderate-sized platform.

Augmentation and Expansion of the NSF Astronomy Mid-Scale Program (Highest Priority Sustaining Activity for Ground)

Mid-scale programs—across the entire range of ~$4 million to 120 million—enable new transformative capabilities by incentivizing creative approaches from the community for cutting-edge instruments and experiments. They also ensure robust capabilities for basic research through continually refreshed instrumentation suites and can respond rapidly to strategic priorities. For these reasons it is essential to expand funding levels for the astronomy funding available through mid-scale programs, MSIP and Mid-scale Research Infrastructure (MSRI). It is also essential to add components to the astronomy mid-scale program to target strategic areas through dedicated calls, and to sustain and advance instrumentation on existing telescopes. For the next 10 years, time-domain astrophysics, highly multiplexed spectroscopy, and radio instrumentation (including radio transient cameras and neutral hydrogen mappers) are the priorities for strategic calls. Dedicated calls are also needed to ensure the regular upgrading of instrumentation on existing facilities, with an emphasis on 4–10 m class

optical/infrared telescopes. These two new elements would be added, in addition to entirely open competitions of new ideas, in a balanced way that responds to proposal pressure.

1.5.4 Programs that Enable Future Visions

Great Observatories Mission and Technology Maturation Program (Highest Priority for Enabling Programs for Space)

NASA's flagship missions are driven by transformative scientific visions, and they advance a broad range of scientific objectives. Overlapping or near-simultaneous wavelength coverage is particularly impactful, as evidenced by the success of NASA's Great Observatories. Given the large costs and development timescales associated with the large missions presented to this survey, achieving this will only be possible if a new approach is taken to mission maturation, and in particular phasing it with decadal survey advice. The Great Observatories Mission and Technology Maturation Program is aimed at increasing the cadence of large missions by designating appropriate scope at an early stage and making significant investments in maturing missions to the appropriate level prior to ultimate recommendation and implementation. This motivates the recommendation that a large IR/O/UV mission first enter the maturation program, and only when that has been successful as defined by a review, would it proceed to formulation. It is also important that additional missions enter the maturation program in the next 10 years to ensure the needed cadence for panchromatic capabilities, and the priorities for this are a far-IR flagship with some of the capabilities of the proposed Origins, and a high-resolution X-ray mission with some of the capabilities of the proposed Lynx. An important aspect is that both the X-ray and far-IR missions are to be matured with cost targets of \$3 billion to \$5 billion. Determining the range of capabilities for these missions will be part of this maturation program, and will be informed by the first Probe mission selection.

Technology Development for Future Gravitational Wave Observatories (Ground)

Gravitational wave astrophysics is one of the most exciting frontiers in science. One of the survey's key priorities is the opening of new windows on the dynamic universe, with gravitational wave detection at the forefront. The continued growth in sensitivity of current-generation facilities, such as LIGO, through phased upgrades and planning the next-generation observatory, such as Cosmic Explorer, is essential. This will require investment in technology development now. The survey committee strongly endorses gravitational wave observations as central to many crucial science objectives. Because the technology development for future upgrades and observatories is funded by NSF Physics, it is beyond the survey's charge to formally recommend this investment.

1.5.5 Large Programs that Forge the Frontiers

A Future Large Infrared/Optical/Ultraviolet Telescope Optimized for Observing Habitable Exoplanets and General Astrophysics (Highest Priority for Space Frontier Missions)

Inspired by the vision of searching for signatures of life on planets outside of our solar system, and by the transformative capability such a telescope would have for a wide range of astrophysics, the priority recommendation in the frontier category for space is a large (~6 m diameter) IR/O/UV telescope with high-contrast (10^{-10}) imaging and spectroscopy. This is an ambitious mission, of a scale comparable to the HST and JWST space telescopes. It is also one that will be revolutionary, and that worldwide only NASA is positioned to lead. A period of mission and technology maturation is required, however with

sufficient investment this could be completed before the end of the decade, and the mission could commence formulation prior to 2030. (Section 7.5.2)

Decision Rules: Prior to commencing mission formulation, a successful Great Observatories Mission and Technology Maturation program must be completed, and a review held to assess plans in light of mission budgetary needs and fiscal realities.

The U.S. Extremely Large Telescope Program (Highest Priority in the Ground-Based Frontier Category)

Because of the transformative potential that large (20–40 m) telescopes with diffraction-limited adaptive optics have for astronomy, and because of the readiness of the projects, the survey committee's top recommendation for frontier ground-based observatories is investment in the U.S. ELT program. The U.S. ELT program is made up of three elements: the Giant Magellan Telescope (GMT), the Thirty Meter Telescope (TMT), and NSF's National Optical-Infrared Astronomy Research Laboratory (NOIRLab). The primary mirror of the GMT has a total diameter of 24.5 m and the telescope has a 25 arcmin field-of-view (FOV). The GMT will be located at the Las Campanas Observatory in Chile. The TMT primary mirror has a diameter of 30 m, and the telescope has a 20 arcmin FOV. The TMT will either be sited on Maunakea in Hawaii, or at Roque de los Muchachos Observatory on La Palma in the Canary Islands. These observatories will create enormous opportunities for scientific progress over the coming decades and well beyond, and they will address nearly every important science question across all three priority science themes. Both projects are essential for keeping the U.S. community's global scientific leadership, providing important synergistic capabilities that complement those planned for the European ELT. However, both projects have significant remaining risks primarily associated with the need to raise additional private or international contributions. The success of at least one of these projects is absolutely essential if the United States is to maintain a position as a leader in ground-based astronomy. The objective is to achieve a time share that is equivalent to 25 percent in each telescope. If only one project is viable, then a larger fraction on that telescope is required to meet the survey's scientific goals, with the aim of achieving an NSF share up to 50 percent time in that project. (Section 7.6.1.1)

Decision Rules: Successful completion of an external review that will determine the financial viability of both projects, final site selection (in the case of TMT), development of an appropriate management plan and governance structure, and appropriate plans for public access and data archiving.

The Cosmic Microwave Background Stage 4 Observatory (CMB-S4)

Given technical and scientific progress over the last decades, ground-based studies of the CMB are poised to take a major step forward in the coming decade. The Cosmic Microwave Background Stage 4 (CMB-S4) observatory will leverage this progress and will have broad impact on both cosmology and astrophysics. Realizing the ultimate scientific potential of ground based CMB observations will take an effort far beyond what can be achieved by independently scaling up existing experiments. CMB-S4 observatory, a joint effort of NSF and DOE, is the compelling and timely next leap for ground-based observations. It will conduct a 7-year ultra-deep survey of a few percent of the sky from the South Pole with a combination of large and multiple small aperture telescopes observing from 30-270 GHz. This will be done in parallel with a 7-year deep/wide survey of roughly half the sky with additional telescopes sited in the Atacama desert in Chile. The Survey is also excited by the breadth of science, including time-domain and transient studies, and the potential engagement of a community well beyond traditional CMB cosmologists. To maximize the science, transient alerts and well calibrated maps from all surveys will need to be made available to the entire community in a timely fashion, even if it requires some extra resources to do so. (Section 7.6.1.3)

The Next Generation Very Large Array (ngVLA)

For the past four decades, the Karl Jansky Very Large Array (JVLA) and the Very Long Baseline Array (VLBA) have been the premiere observatories worldwide for accessing the sky at centimeter wavelengths. It is of essential importance to astronomy that the JVLA and VLBA be replaced by an observatory that can achieve roughly an order of magnitude improvement in sensitivity compared to these facilities, with the ability to image radio sources at centimeter to millimeter wavelengths on scales of arcminutes to fractions of a milliarcsecond. The ngVLA is such a facility; however, it is immature in its development, and considerable effort must be put into studies to understand and reduce the cost relative to current estimates, secure international partnerships, and prototype the antennae. With such an effort commencing soon, the ngVLA would be ready to commence construction by about 2030. It will be important to begin implementation as soon as it is technically and fiscally possible. (Section 7.6.1.4)

Decision Rules: Implementation is contingent on a successful design, development and prototyping program, cost studies, and commitments from any foreign partners. A review will determine the project's readiness and consistency with budgetary constraints prior to commencement of construction.

The IceCube-Generation 2 (IceCube-Gen2) Neutrino Observatory

Observations of high-energy neutrinos enable astrophysical advances in the study of some of the most energetic phenomena in the universe. The IceCube-Gen2 would greatly enhance the capabilities relative to IceCube, would be able to resolve the bright, hard-spectrum TeV-PeV diffuse neutrino background into discrete sources, and would make the first detections at higher neutrino energies. Multi-messenger astrophysics is a major theme of this report, and the survey endorses the IceCube-Gen2 observatory as important to many key survey scientific objectives. Because it is funded by NSF Physics, it is beyond the survey's charge to recommend this investment. (Section 7.6.2.3)

1.6 ADDITIONAL ADVICE

In addition to the vision for new, recommended future endeavors, this decadal survey report offers advice on aspects of the agencies' programs aimed at optimizing returns for their existing programs.

Data Archives. An important component of creating effective archives is coordinating with cross-agency and international archiving services to develop best practices and interoperability. As a step toward this, it is important for NASA and NSF to explore mechanisms to improve coordination among U.S. archive centers and to create a centralized nexus for interacting with the international archive communities. The goals of this effort are best defined by the broad scientific needs of the astronomical community.

Solar Physics. Solar physics is directly relevant to astronomy, as it is the study of our nearest star, and interacts with stellar astrophysics; is input to studying the Earth-Sun connection and expanding to stellar-planetary interactions; and is vital to understanding Earth's climate and space weather. The survey committee concluded that an appropriate role for astronomy and astrophysics decadal surveys is to comment on the value of ground-based solar physics projects for astronomy and astrophysics priorities, with the solar and space physics decadal survey being the more appropriate body to prioritize and rank ground-based solar physics projects within the context of the full range of multi-agency activities in solar physics.

NSF Portfolio Reviews. Regular reviews of more mature observatories are essential to determine how to optimize their scientific return and cost effectiveness, and to determine when a facility is at the end of its operational life. While some aspects of ground-based facility reviews are considered as part of the review of operating agreements for observatories, these are not an appropriate substitute for a review that considers the entire portfolio on a self-consistent, holistic basis. It is essential that NSF AST establish a regular cadence of reviews of its operational portfolio, at a frequency sufficient to respond to changes in scientific and strategic priorities in the field. An appropriate target is two reviews per decade.

SOFIA. The survey committee has significant concerns about SOFIA, given its high cost and modest scientific productivity. The NASA portion of SOFIA's operating budget is out of balance with its scientific output, which is a fraction of that of comparable cost missions (e.g., HST, Chandra) and often less than those of Explorer missions. The survey committee finds no evidence that SOFIA could transition to a significantly more productive future and notes the minimal mention of SOFIA science by the science panels. The committee found no path by which SOFIA can significantly increase its scientific output to a degree that is commensurate with its cost and endorses NASA's current plan to discontinue operations in 2023.

NASA's Balloon Program. NASA's balloon program plays an important role in offers access to a near-space environment with a wide variety of options for duration and sky coverage, for developing technologies, and training future generations of technologies and mission leaders. It is, however, clear that the balloon program is not yet achieving the potential promised by the advent of ultra-long duration balloon (ULDB) flight capabilities. It is important that the balloon program be critically reviewed to evaluate how to optimally support innovative payload development and to increase the cadence and reliability of LDB and ULDB flights.

NASA's Program of Record. NASA's upcoming Roman Space Telescope, and ESA's Athena X-ray Observatory and LISA mission, in which NASA is a significant partner, are essential to the survey's science program. Advice on how to optimize the science return includes: holding a non-advocate review of Roman Space Telescope's science program to set the appropriate mix of survey time to guest investigator-led observing programs; and at the appropriate time, establishing funding for LISA science at a level that ensures U.S. scientists can fully participate in LISA analysis, interpretation, and theory.

1.7 CONCLUSION

The integrated program forwarded in this report advances a vision for discovery and progress for the coming decades. The content of the remaining chapters, together with the panel reports, represent an enormous effort that took years of preparation on the part of a large fraction of the astronomical community, and more than 2 years for the survey and its committees to complete. The full context of the recommendations and advice summarized in this chapter can only be appreciated by reading the report in its entirety. Realizing the opportunities presented in these pages will only be possible with the continued dedication and energy of the community, the agencies, and the excitement of the nation to explore the cosmos and answer some of humanity's most profound questions.

2

A New Cosmic Perspective

The past decade has been one of extraordinary discoveries in astronomy and astrophysics, and a realization of the scientific vision of the 2010 decadal survey, *New Worlds, New Horizons in Astronomy and Astrophysics.*[1] Scientific advances range from the detection of gravitational waves from merging black holes, to a direct image of a black hole in a nearby galaxy, to the production of heavy elements in the merger of two neutron stars, long-hypothesized but observed in exquisite detail for the first time. Increasingly sensitive instrumentation and powerful simulations are uncovering connections between the complex gaseous surroundings of galaxies and the forces that shape them. An explosion in the number of known exoplanetary systems has been accompanied by the detailed characterization of a subset of these other worlds, with insights into their formation arising from imaging of the disks where young planets are assembling. Newly discovered fossil structures from the formation of the Milky Way Galaxy open a window on the Milky Way's distant past, and observations take several steps closer to diagnosing the conditions present shortly after the Big Bang.

The investments of previous decades bore fruit in this decade in the awarding of Nobel Prizes in Physics for six discoveries derived from astronomical measurements: dark energy, neutrino oscillations, gravitational waves, exoplanets, physical cosmology, and black holes (Figure 2.1). At the same time, international collaborations greatly expanded. A salient example of this is the seminal paper detailing the discovery of two merging neutron stars and their gravitational and electromagnetic signatures: it encompassed nearly 4,000 authors from 900 institutions and 70 observatories, spanning all seven continents and space-based facilities, or roughly one third of the professional astronomical community as well as most of the gravitational wave community world-wide.

[1] National Research Council, 2010, *New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/12951.

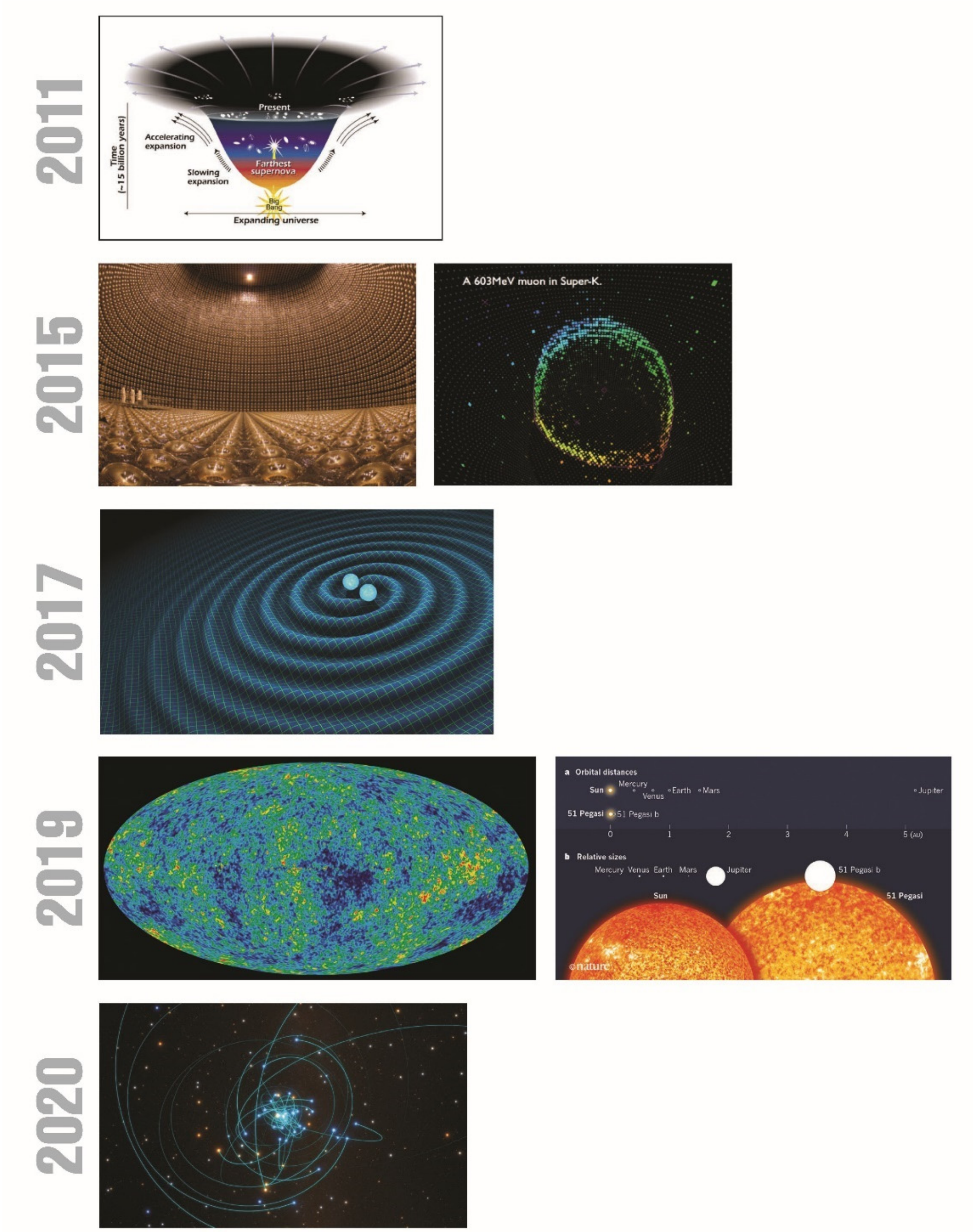


FIGURE 2.1 Physics Nobel Prizes derived from astrophysical measurements. The decade of 2011-2020 witness the awarding of Nobel Prizes for six different astronomical discoveries. In 2011, Saul Perlmutter, Adam Riess, and Brian Schmidt received the prize "for the discovery of the accelerating expansion of the universe through observations of distant supernovae." The citation to Takaaki Kajita and Arthur McDonald in 2015 was "for the discovery of neutrino oscillations, which shows that neutrinos have mass." In 2017, Kip Thorne, Rainer Weiss and Barry Barish were awarded the prize "for decisive contributions to the LIGO detector and the observations of gravitational waves." The year 2019 saw the awarding of the Nobel Prize in Physics to James Peebles "for theoretical discoveries in physical cosmology" and to Dider Queloz and Michael Mayor "for the discovery of an exoplanet orbiting a solar-type star." Most recently, in 2020, the topic of black holes received Nobel attention, with recognition to Roger Penrose "for the discovery that black hole formation is a robust theory of general relativity" and to Andrea Ghez and Reinhard Genzel "for the discovery of a supermassive compact object at the center of our galaxy." SOURCE: *2011:* NASA/STScI/Ann Field; *2015:* Kamioka Observatory, ICRR (Institute for Cosmic Ray Research), The University of Tokyo; *2017:* R. Hurt/Caltech-JPL; *2019-left:* NASA/WMAP Science Team; *2019-right: pending*; *2020:* ESO/L. Calcada/spaceengine.org.

The *New Worlds, New Horizons* decadal survey identified three main science objectives for their decade, while also enabling a wider discovery potential. The resulting scientific program is still bearing fruit in this decade and will in the next. On the topic of *Cosmic Dawn: Searching for the First Stars, Galaxies, Black Holes*, the James Webb Space Telescope (JWST) (launch expected in 2021) will directly examine the youngest observable galaxies, the Vera Rubin Observatory (science first light expected in 2021) and Nancy Grace Roman Observatory (launch expected in 2025) will transform views of dwarf galaxies at the extremes of galaxy formation and the record of ancient stars they left behind, and the European Space Agency's (ESA) Laser Interferometer Space Antenna (LISA) (launch expected in mid-2030s) should identify mergers of black holes all the way back to their earliest formation. In the arena of *New Worlds: Seeking Nearby Habitable Planets*, the Transiting Exoplanet Survey Satellite (TESS) (launched 2018), JWST, and soon Roman will explore a widening array of exoplanet types. For the *Physics of the Universe: Understanding Scientific Principles* objective, myriad ground- and space-based telescopes will use the universe as a laboratory to probe the nature of dark matter and dark energy. Roman, Rubin, ESA's Euclid (launch expected in the latter half of 2022), the most recent Laser Interferometer Gravitational-Wave Observatory (LIGO) upgrade to Advanced LIGO Plus (operations expected to begin 2024), the Event Horizon Telescope, and ground cosmic microwave background (CMB) experiments will advance understanding of the conditions present in the infant universe, the properties of dark energy, and the fundamental physics associated with black holes.

The development of scientific priorities for this survey began with the receipt of 573 science white papers (written by 4516 unique authors and/or endorsers) in early 2019; 573 such papers were received. These papers formed the starting point for deliberations by six expert science panels, organized by subfield: Compact Objects and Energetic Phenomena; Cosmology; Galaxies; Exoplanets, Astrobiology and the Solar System; The Interstellar Medium and Star and Planet Formation; and Stars, the Sun, and Stellar Populations. Each panel reviewed recent progress in their fields and identified key challenges and priorities for the coming decade and beyond. Their reports are included as an Appendix to this report. From these wide array of opportunities each panel identified four key science questions regarded as being especially ripe for investigation in the coming decade, along with a "discovery area" where emerging capabilities or techniques offer great promise for major advances. Tables 2.1 and 2.2 at the end of this chapter provide a summary list of these questions and discovery areas. These are not intended to encompass all of the important science needed, but rather to highlight particularly important questions and opportunities.

The panel reports were then integrated by the steering committee into the summary science case which forms the remainder of this chapter. It soon became clear that most of the science questions and discovery areas could be organized into three broad thematic areas: *Worlds and Suns in Context* highlights the extraordinary advances over the past decade in the study of exoplanets, stars, and their associated planetary systems, and the opportunities for transformational advances in these areas, including the ultimate search and characterization of habitable planets, in the decades ahead. *Cosmic Ecosystems* represents an integration and culmination of understanding the origins of galaxies, stars, planets, and massive black holes, and the realization that the life cycles of the universe over this billionfold range of scales are intimately connected, through feedback processes propagating through the gas within, surrounding, and between galaxies. The *New Messengers and New Physics* theme embodies the dual revolutions brought about by the marriage of observations of light with those from gravitational waves and elementary particles (multi-messenger astrophysics) along with the expansion of measurements of the sky over time (time-domain), as well as the opportunities for major advances answering some of the most fundamental questions in astrophysics and physics, such as the nature of dark matter and dark energy. Each of these three themes is summarized below, and Tables 2.1 and 2.2 list the thematic areas associated with each of the science panel questions and discovery areas. All are represented in the themes and many cross multiple themes.

Although the three overarching themes effectively distill the multitude of science goals and priorities contained within the six panel reports, their encompassing nature do not provide clear scientific

goals which can be easily grasps by the non-professional readers of this report. With that in mind we have identified a key priority science question for each of the themes, which help to motivate the strategic program later in this report: Pathways to Habitable Planets; New Windows on the Dynamic Universe; and Unveiling the Drivers of Galaxy Growth. The priority areas for each theme are described at the end of each respective section.

2.1 WORLDS AND SUNS IN CONTEXT

The planets in our solar system, and the Sun at the center of it, provide the most direct connection to the myriad other stars and planets in our galaxy and the universe. With the flowering of capabilities expected in the next decade, progress in both stellar astrophysics and planetary science will expand and provide a broader context with which to understand and appreciate our cosmic perspective.

Within the last decade, progress in planetary demographics has achieved a reversal of sorts; new knowledge that planets are common, along with multi-planet systems, translates to a new cosmic understanding that there are likely more planets than stars in the universe. The Copernican revolution continues, in the realization that amongst the incredible variety of exoplanetary systems, our own solar system may prove to be an outlier rather than an exemplar. Along with pondering the cosmic order to understand stars, their formation, evolution, and ultimate fate, there is a parallel track for planets—how do they form and evolve?—and a merging of these tracks when considering questions related to habitability. It is an exciting time in which to practice the astronomical craft, as humanity edges ever closer to being able to answer the age-old question “Are we alone?” It is humbling and exciting to contemplate that the question of whether life exists elsewhere could be answered with the technology humanity now possesses.

The past decade has also witnessed a renaissance in stellar astronomy. Gaia, an ESA mission to deliver fundamental stellar parameters such as distances and three-dimensional motions on the sky, has proved revolutionary even with its initial data releases, to articulate the connections between and among groups of stars. Other time-domain observatories primarily designed to detect exoplanets have returned a wealth of asteroseismological observations allowing us to probe the interior structures of stars. Whereas in the past only a few fundamental parameters of a star could be determined accurately, and most of these in a relative sense, now mass, size, distance, age and chemical makeup for a wide swath of stars are measurable. The coming decade will continue this trend: where Gaia enabled precision stellar parameters of roughly a billion stars in the Milky Way, the Rubin Observatory promises to expand by a factor of 10 the number of main sequence stars for which distances can be determined.

2.1.1 Stellar Demographics

We are in the midst of a stellar renaissance, as astronomers come to know the individual journey of each star, separating them from anonymous points of light in the heavens to having the equivalent of detailed dossiers of physical characteristics and histories.

FIGURE 2.2 Mosaic image of part of our neighboring galaxy, Andromeda, as viewed by the Hubble Space Telescope as part of the Panchromatic Hubble Andromeda Treasury Project (PHAT). Roughly 100 million stars in this galaxy's pancake-shaped disk are resolved with this data. Precision characterization of the individual stellar properties enables determination of factors affecting galactic structure and evolution. SOURCE: https://hubblesite.org/image/3476/gallery/73-phat; NASA, ESA, J. Dalcanton, B.F. Williams, and L.C. Johnson (University of Washington), the PHAT team, and R. Gendler

This increasing complexity of stellar astrophysics extends to exposing the internal states of stars and the insights that come along with that revelation. Asteroseismology was a once-bespoke method of sounding the internal state of a star via detections of oscillation modes. The technique is now an established tool to determine precise stellar ages in large data sets, as well as returning masses and radii which can be used to calibrate models. The emergence of ultra-precise, long-duration, and continuous light curves from space (first with the European mission Convection Rotation and planetary Transits [CoRoT], then with NASA's Kepler and TESS as well as ESA's Characterizing Exoplanets Satellite [CHEOPS] and soon, ESA's Planetary Transits and Oscillations of stars [PLATO] mission) served both the exoplanet community and the stellar astrophysics community by enabling planet detections via the transit method and vastly expanding the number of stars for which asteroseismic oscillations (and subsequent stellar astrophysics studies; see below) could be detected. Probes of the internal states of stars via this method now return constraints on stellar structure previously only theorized. The cores of red giant stars appear to rotate faster than the surface, and oscillation frequencies differentiate red giant stars in which core helium burning is occurring, versus those only burning hydrogen in a shell. Latitudinal differential rotation in the convection zones of Sun-like stars, revealed through asteroseismic observations, indicate a shear much larger than predicted from numerical simulations. The mass distribution of red giant stars probed by asteroseismology does not match predictions from stellar population models. Asteroseismology detects strong magnetic fields in the cores of red giant stars. Indeed, high-precision, high-cadence light curve observations now join traditional methods of photometry and spectroscopy as essential tools of observational stellar astrophysics.

The next decade will continue to provide precision tests of stellar evolutionary models, and ultimately advance our understanding of the structure, dynamics, and evolution of galaxies as a whole (Figure 2.2). The elements present in a star's atmosphere reveals its origins and refines the understanding not just of that star, but when applied to large samples of stars, reveals how the assemblage of stars in our galaxy came to have its form, structure and content. Identifying and studying particular stellar subsamples such as cool subdwarfs provide the fossil record of the early history of star formation in the Milky Way Galaxy, as their elemental compositions are unchanged from that at their birth. Necessary links in this chain of chemical tagging are improvement in the laboratory and numerical calculations of atomic and molecular transitions and opacities, together with inclusion of more realistic geometries beyond assuming stars are spherical, and non-equilibrium effects in stellar model atmospheres. In the coming decades, high-resolution spectroscopy with extremely large telescopes will expand these abundance measurements

throughout the Local Group. Connecting this chemical record of the galaxy with the dynamical record from Gaia and spectroscopy should produce a definitive fossil record of the assembly and life history of our galaxy. Progress in the next decade will require advances in industrial-scale spectroscopy (e.g., SDSS V and future even larger surveys) along with computational methods to harness the large data sets of photometric, spectroscopic and astrometric surveys.

Whether a star has one or more partners, and the nature of those partners, influences its life history because of mass exchange, particularly for massive stars. This is especially true for the end state of stellar evolution, as the formation of compact objects provides connection points to probes of extreme gravity and extreme particle acceleration. The detection of gravitational waves from astrophysical objects this decade leads to invigorated research into the endpoints of stellar evolution for the next decade. Stars in multiple systems can have very different evolutionary pathways compared to their single counterparts; detailed photometric and spectroscopic electromagnetic observations spanning infrared through X-ray wavelengths, coupled with gravitational wave measurements and attention from theory, will elucidate these multiple routes and their consequences. Theoretical modelling of binary co-evolution is necessary to understand the fate of close binaries that do interact. White dwarfs caught in the act of merging will be studied by the Rubin Observatory, and can be linked to resolved gravitational wave signals detected by LISA once it launches, as these systems will also likely contribute a major part of the stochastic background expected from the Milky Way Galaxy. Further observational and theoretical research will sharpen constraints on the mass threshold that separates an end state of a white dwarf from core-collapse supernova formation. The effects of rotation, binary, and magnetic fields can be implemented in population synthesis models, with wide applicability ranging from understanding the ionizing output of massive stars to interpreting the spectra of distant galaxies. More generally, mapping out the myriad evolutionary pathways by which stars come to populate every part of the Hertzsprung-Russell diagram will require a robust observational and theoretical understanding of the formation, evolution, and especially interaction of stars in multi-star systems.

The angular momentum of a star, typically measured by its surface rotation period, is a key parameter to unlocking its current state and is a fundamental parameter in its own right. Determining rotation periods for a few tens of stars would in years past have been the subject of a Ph.D. dissertation. Now, with automated light curve analysis of stellar targets to search for transiting planets, a click of a button is practically all that is needed to compute rotation periods of tens of thousands of stars within a single research paper. This has proceeded in recent years largely as the stellar byproduct of exoplanet studies (Figure 2.3), a fortuitous result but one biased by the particular target selection criteria used for exoplanet searches. The next decade will see rotation rates of stars (and other time-domain stellar astrophysics measures such as flaring) determined over nearly the entire galaxy, likely with the aid of modern machine learning practices. Application of these rotation periods as clocks for stellar age dating converts stellar time-domain astronomy into the industrial-scale extraction of stellar birth dates. This furthers the personalization of stellar histories and adds precision to testing stellar models, as well as providing ground-truth age dating to constrain models of the Milky Way's formation and evolution history.

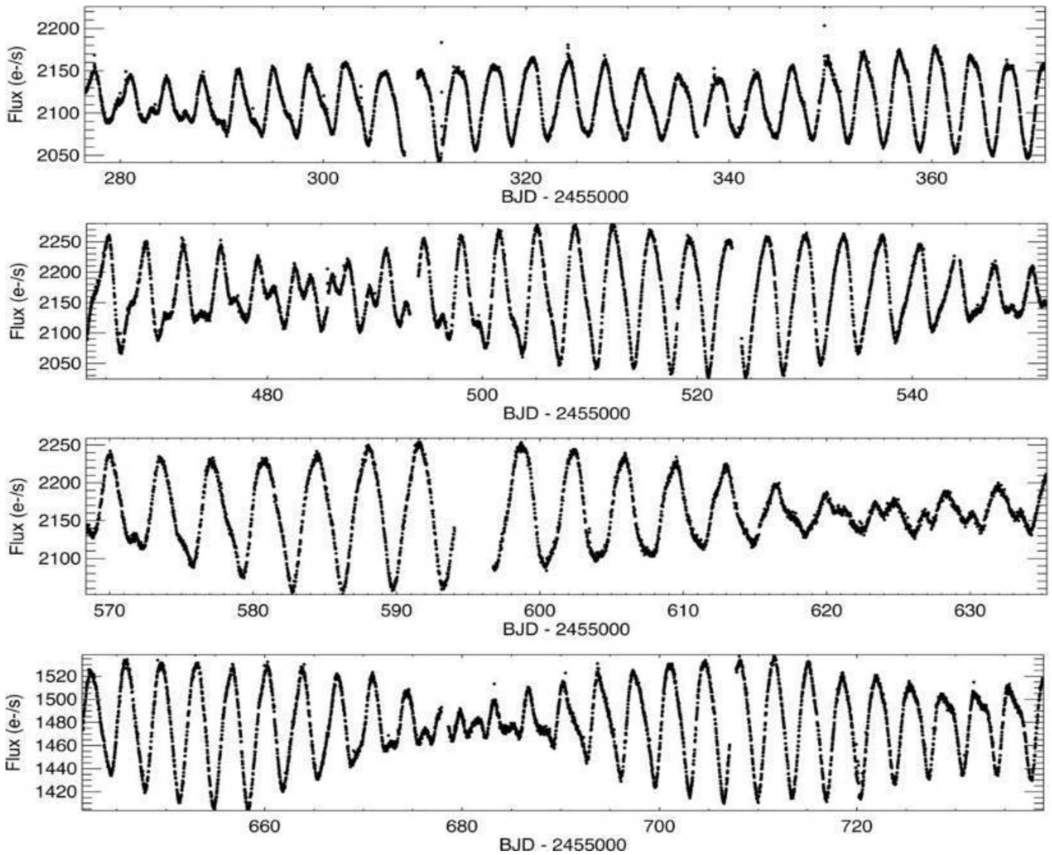


FIGURE 2.3 Evidence of stellar rotation appears from periodic surface features in the Kepler light curve of this K-type dwarf studied by Roettenbacher et al. (2013). Long-term evolution in the number and distribution of starspots is indicated by the changing patterns in the light curve, as well as short-term brightenings that indicate flares. SOURCE: Adapted from "Imaging starspot evolution on Kepler target KIC 5110407 using light curve inversion," Rachael M. Roettenbacher et al 2013 *The Astrophysical Journal* 767 60, © AAS. Reproduced with permission. doi:10.1088/0004-637X/767/1/60.

Amongst all the detailed stellar knowledge, astronomy still discovers entirely new types of stars and stellar phenomena. Progress in the next decade is expected to reveal more and different types of stellar exotica, as testaments to the incompleteness of the knowledge about the many forms and varieties of stellar behavior. The unprecedented dimming of Betelgeuse**,** one of the brightest stars in the sky and the shoulder of the Orion constellation, from December 2019 through May 2020, proved that even well-studied stars can throw an unexpected astronomical curveball. Multi-wavelength observational capabilities which include ultraviolet and even X-ray wavelengths provide key diagnostics of temperature, density, abundances, and kinematics, essential for the post-mortem understanding of these events. The rare phenomena, such as the unusual dimmings of Boyajian's star, sparsely erupting pre-main sequence stars, or the elusive luminous blue variables, will become commonplace, and understanding will hopefully follow discovery of stars and their environments in a virtuous cycle.

The Sun is a singular star amongst all the others, primarily for its proximity but also because of our dependence on its behavior for our existence. Observations of the Sun are necessarily a touchstone for virtually all of stellar theory, with ripples throughout the understanding of all of stellar astrophysics. The anticipated first science to be done with the revolutionary Daniel K. Inouye Solar Telescope (DKIST) facility promises more answers to fundamental questions of how magnetism controls our star. This four-meter, solar-dedicated telescope on the ground (a "Hubble for the Sun"; see Figure 2.4) represents a large leap in solar observing capabilities. The interplay between magnetic flux and mass flows is of universal

astrophysical importance. This decade's detailed observations will have the requisite spatial, temporal, and spectral resolution as well as the dynamic range to provide a ground-truth for models of basic magnetic structures. Advances originating from this zoomed-in view of solar magnetic structures require complementary global measurements—particularly of the solar corona—to understand magnetic energy storage and release. Complementary radio observations could generate three dimensional mapping of the magnetic field in sunspots, and magnetic field measurements in the global corona provide context for the more detailed, restricted field of view measurements.

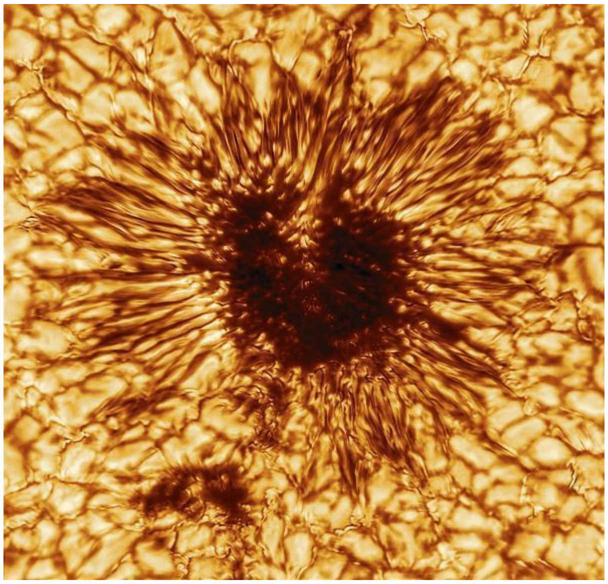


FIGURE 2.4 This is the first sunspot image taken on January 28, 2020, by the National Science Foundation's Daniel K. Inouye Solar Telescope's (DKIST) Wave Front Correction Context Viewer. The image reveals striking details of the sunspot's structure as seen at the Sun's surface. The sunspot is sculpted by a convergence of intense magnetic fields and hot gas roiling up from below. This image uses a warm palette of red and orange, but the context viewer took this sunspot image at the wavelength of 530 nanometers - in the greenish-yellow part of the visible spectrum. SOURCE: NSO/AURA/NSF (see https://nso.edu/press-release/inouye-solar-telescope-releases-first-image-of-a-sunspot/).

The singular nature of the Sun is both its blessing and its curse. Its ability to provide constraints on astrophysical questions with unrivalled accuracy is in tension with its fitness as a prototype for all things stellar. While it is a spectacularly well-studied star, it is a unique case study observed at one point in its 4.6 billion year evolutionary history. Observations in the last decade provided tantalizing hints that the Sun's cycle of magnetic activity does not behave the same as other solar-like stars. Work in the next decade will broaden the range to include a larger sample and provide the perspective of stars of different ages and spectral types. While hints exist now of a disconnect in the relationship between fundamental stellar parameters and magnetic properties—magnetically active stars can have larger radii than predicted based on their temperatures; stellar twins identical in all other properties can have very differing levels and amounts of magnetic field distribution and magnetic activity signatures—observations in the next decade will further this discovery space and fill important gaps in the foundation of stellar structure theory.

On a broader scale, it is clear that stars are not uniform discs of light. Time-domain astronomy enables the characterization of surface structure using the changes in light seen either in broad wavelength bands (Figure 2.3) or in narrow spectral features that reveal inhomogeneities. Optical and infrared interferometric observations have begun to localize these regions. Observations spanning multiple regions of the electromagnetic spectrum provide independent constraints on the properties of these features,

characterizing the magnetic, chemical, or cloud-like nature of the features. Indeed, the ability to measure rotation periods arises from the periodic brightening and dimming caused by blemishes passing across the face of each star. On stars like the Sun those spots reveal the telltale existence of subsurface magnetism; the multitude of long-duration high precision photometric light curves available from space missions including CoRoT, Kepler, and TESS, reveal the ubiquity and variety of surface inhomogeneities of cool stars. Future progress in interferometers, from optical to infrared to radio wavelengths, will advance knowledge of the underlying physics causing spatial inhomogeneities on stellar surfaces.

The importance of magnetic fields in a stellar context was expected but is now unequivocal. Magnetism in massive stars has been confirmed (but not entirely explained) in a minority of cases, and stronger magnetic fields than initially expected have been found at the end of the main sequence, in the ultracool dwarfs spanning stellar and substellar origins. A variety of observational tools enable these discoveries: high-energy observations detail the magnetic shaping of stellar winds and unveil the existence of heated plasma in an above-the-surface corona; optical and infrared spectra reveal the tell-tale signature of magnetic splitting of atomic and molecular lines; and radio observations expose the action of accelerated electrons thereby revealing the presence of magnetic fields in their atmospheres. Stellar coronae are common in cool main-sequence stars, but the origin of this heated plasma is still unknown, even on the Sun. Future sensitive high energy and radio measurements will probe the coronae of stars in more detail and expand the number of objects amenable to study. Ultraviolet spectra probe the chromosphere and above, regions in a cool star's atmosphere where magnetic forces begin to dominate. These observational advances also require associated advances in modelling the intricacies of stellar structure for full understanding of the impact of convection, rotation, and magnetic field generation on the complex and dynamic nature of stars.

2.1.2 Exploring Alien Worlds

Over the past two decades an incredible diversity of worlds and systems has been discovered, from giant planets over 10 times the mass of Jupiter to the Trappist-1 system with seven Earth-sized planets packed in an area smaller than the orbit of Mercury. This revolution is ongoing. New capabilities led to an explosion of discoveries in the last decade (Figure 2.5), filling in demographics of planets, expanding the ability to characterize the composition and nature of individual members, and advances in new facilities adding to the depth of characterization possibilities.

These discoveries are helping us understand fundamental questions about ourselves. How did the solar system form? Are systems like our own common or rare? Are planets like Earth common or rare? And, ultimately, do any of those Earth-like planets harbor life? With current technology, the planets in solar systems like our own are nearly undetectable; with new facilities over the next two decades, this will change, and the picture of other worlds will start to become complete.

Finding exoplanets is challenging. Many exoplanet detection methods have been developed, and each gives some important, but limited, information. The radial velocity (RV) technique measures orbits and constrains the mass of a planet using the motion of its parent star. This technique has been used for a number of exoplanet "firsts" over 25 years (including the Nobel Prize winning discovery of 51 Peg b), and astronomers are pushing the technology to be able to detect Earth-mass planets on periods of months, to perhaps a year, in search of objects similar to Earth. In the last decade, another technique, exoplanet transits, has transformed the view of exoplanetary systems, in particular when employed via high precision space missions like NASA's Kepler and TESS missions, and before them the European CoRoT mission. A transit, when a planet passes directly in front of its parent star, allows for a measurement of a planet's radius and orbital period. Kepler showed that planets on close-in orbits (within 100 days) are extremely common (Figure 2.5), which has revolutionized the understanding of planet formation, and is one of many examples that show that architecture of planetary systems is exceedingly diverse.

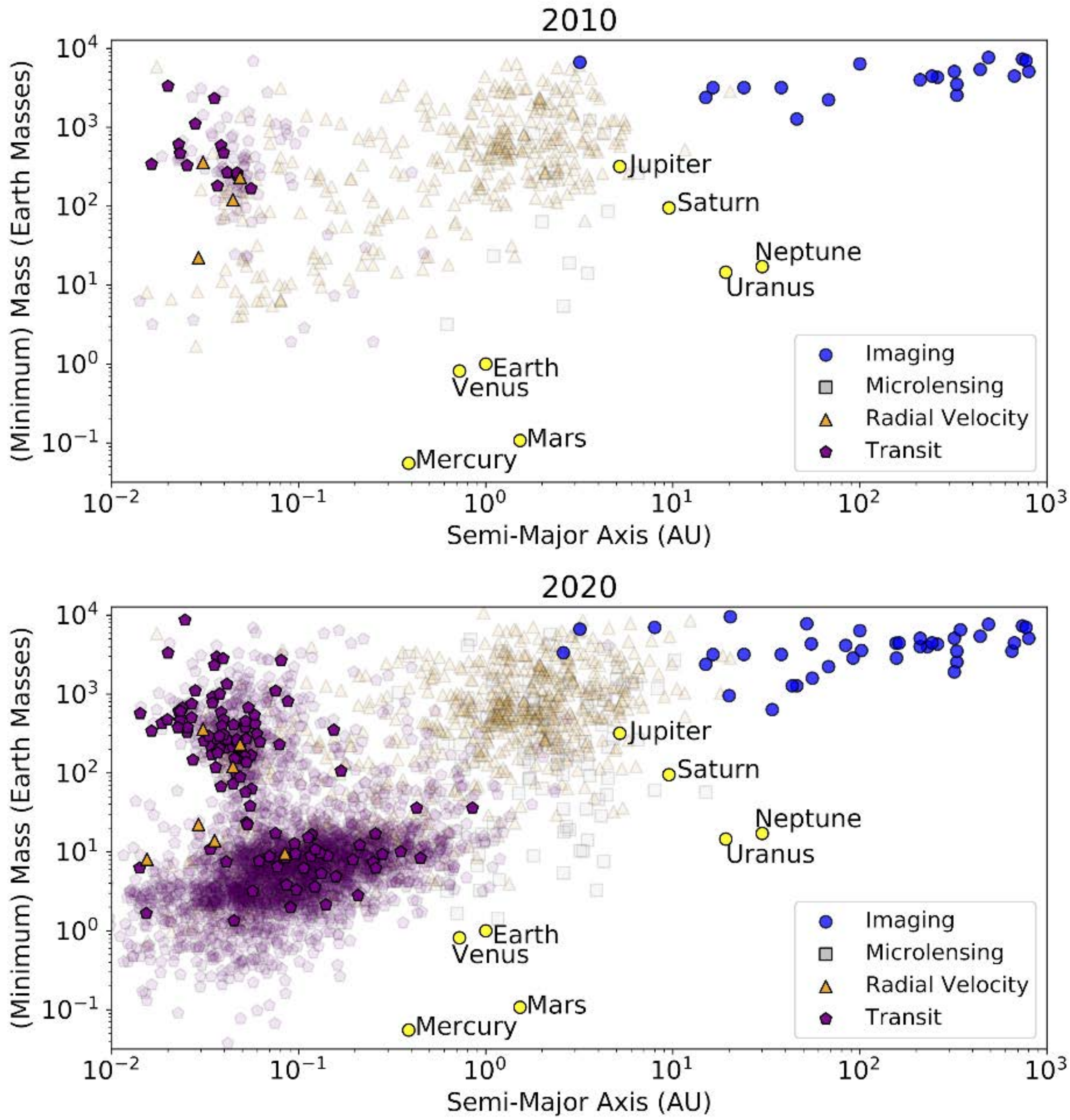


FIGURE 2.5 The population of known exoplanets in 2010 (*top*) and 2020 (*bottom*). Each symbol represents a known extrasolar planet, colored by initial discovery method. Hollow symbols are planets that have been discovered. Filled symbols are planets whose atmospheric composition have been *characterized* by measurements of its spectrum or brightness. Over the past decade astronomers have begun to move from the era of planetary census-taking to detailed characterization, and the next decades will both complete the missing parts of the census—planets like our own solar system—and see an explosion in characterization. SOURCE: D. Savransky and B. Macintosh, with data from the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

The radial velocity and transit methods are best suited for finding planets on relatively close-in orbits. A method prioritized by *New Worlds, New Horizons,* gravitational microlensing, will be used by the Roman Space Telescope in this decade to complete the planetary census by finding planets from 1 to 100 AU, and even free-floating planets. Microlensing exploits the bending of light by the gravity of the

star and planet to detect distant systems. This technique will be powerful for measuring how common such planets are on these wider orbits. However, these planets cannot be followed up for future in-depth observations, because this chance star-planet alignment does not repeat.

Another major technique of long-term importance to the field is direct imaging, where the light of the parent star is blocked to make a faint planet directly visible. Currently this is only practical for young giant planets in the outer parts of solar systems. However, with suitable technology development there is now a clear path forward to use direct imaging techniques on ground-based extremely large telescopes (ELTs) and optical interferometry with 8 m-class telescopes to study the atmospheres of temperate rocky planets around low-mass stars, and to use future space missions to study potentially Earth-like planets around Sun-like stars, as well as the enormous diversity of non-Earthlike planets.

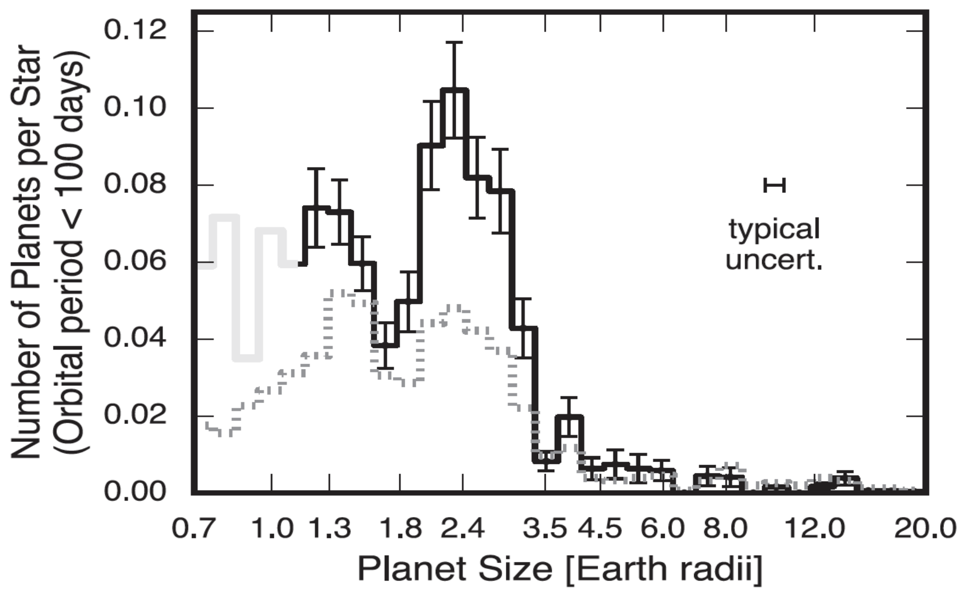


FIGURE 2.6 Plot of number of planets per star vs. planetary radius, for orbital periods less than 100 days, shown in solid black. The dashed grey line is from planet detections before including the Kepler Mission completeness corrections, which becomes more pronounced at smaller planet sizes. The analysis of data from the Kepler Mission showed a pronounced gap between smaller, denser worlds (1-1.7 Earth radii) and larger more Neptune-like planets (1.7-4 Earth radii) with thick hydrogen atmospheres. Detailed analysis of the gap as a function of orbital period and system age suggests that the smaller planets are likely the “stripped cores” of formerly more Neptune-like worlds. SOURCE: Adapted from B.J. Fulton and E.A. Petigura, 2018, “The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap,” *The Astronomical Journal* 156 264, © AAS. Reproduced with permission. doi:10.3847/1538-3881/aae828.

With these diverse techniques astronomers have produced a partial census of exoplanets. The results have been extraordinary, showing that on the average there are at least two planets per star in the Milky Way, and that many planetary systems are very different than our own, with crowded systems of planets intermediate between Earth and Neptune in size. So far it appears that lower-mass M-dwarf stars have more planets than Sun-like stars. Careful characterization of the stars themselves has been crucial to interpreting Kepler results, and broader studies will reveal how these distinctions persist across different regimes. The census remains incomplete; all current planet detection surveys would be essentially blind to every planet in the solar system at their known orbital separations, except Jupiter.

The Kepler Mission has also shown that the changing nature of planets—their time evolution—is also fundamental to understanding the planetary population. Figure 2.6 shows a recent discovery using Kepler data, of a fundamental divide between rocky planets below about 1.7 times the radius of Earth, and larger “sub-Neptunes” that have thick hydrogen-dominated envelopes that increase the planetary radius.

The “gap” between these two populations is thought to be due to the gradual loss of hydrogen-envelopes possibly due to a stellar wind that drives these atmospheres off the planet and into space. The detection of this gap was made possible due to exquisite characterization of the host stars and shows the important connection between stellar and planetary physics.

Approximately twenty Earth-sized (<1.7 R_Earth) planets have been discovered within the habitable zone of low-mass stars—neither too close nor too far, but receiving enough energy to allow liquid water on their surface. A key question that Kepler and other projects have tried to address is the frequency of potentially habitable planets - the average number of Earth-sized planets within the habitable zone of their star, particularly around Sun-like stars. A decade ago, this quantity—known as η_{Earth}—was almost completely uncertain. These planets are very difficult to detect directly, so their exact occurrence rate must be extrapolated from planets higher in mass, closer to their star, and/or from the frequency of habitable zone planets orbiting lower-mass stars. Current constraints indicate that 18-28 percent of Sun-like stars have an Earth-sized (80 to 140 percent of Earth’s radius) planet in their habitable zone, enough to make their study with a future large mission practical. (The number is likely a factor of two higher for cooler lower-mass stars.) The nature of these planets is of course almost completely unknown, whether they have followed a path of planetary evolution amenable to life or something completely different than Earth; but that very uncertainty is a compelling scientific reason to try and answer this most profound question.

2.1.2.1 Planet Formation

The detailed process of planet formation is one of the great unknown frontiers in astrophysics. The core concept of gas and tiny dust particles in a circumstellar disk assembling into planets is well known. But the details of how particles assemble, how the process overcomes barriers to operate quickly enough before the disk dissipates, how this leads to the incredible diversity of known planets, and whether this process can frequently produce planets with Earth’s key characteristics, remain unknown. Understanding this is crucial to placing potentially habitable planets in context, including questions as to how water and atmospheric volatiles are delivered to Earth, whether giant planets affect the evolution of terrestrial worlds, and whether “mini-Neptune” planets can evolve towards habitability. More sensitive observations of planet-forming disks to understand the astrochemistry, dynamics, and role of water in the formation of habitable planets through radio, mm, and far IR spectroscopy will help advance this field.

Planetary demographics and composition only reveal the endpoint of this process. Stunning observational advances from ALMA show complex structures in planet-forming disks (see Figure 2.20), and recent advances with large ground-based telescopes have even caught a protoplanet with its own accretion disk (Figure 2.7). These structures, however, still correspond to large planets in the outer parts of solar systems; in the habitable zone or the regime of the Kepler super-Earths the process itself remains almost inaccessible. Current observations of continuum dust emission cannot probe the innermost regions of the planet-forming disks where Earth-like planets may be located; sensitive radio interferometry at longer wavelengths than those probed by ALMA is needed to see the optically thin emission originating in these locations. Spectra at far infrared wavelengths would provide a unique and revolutionary census of water within these disks, which is a key to understanding giant planet formation and the distribution of water among terrestrial planets. New radio and far infrared surveys of planet forming disks would enable leaps in understanding that would surpass this era’s ALMA-driven revolution.

There have been tremendous advances in the theory that underlies planet formation, as new ideas such as the “streaming instability” and “pebble accretion” use gas and particle interplay and physical changes in the disk to trap planet-forming material and rapidly form giant planets. The models incorporate uncertainties such as turbulence and feedback between the forming planets and disks. Better observations of gas and dust in disks (especially on smaller scales and across a range of ages) and of forming planets (especially of lower masses) will advance this field, combined with larger-scale computational models and laboratory experiments.

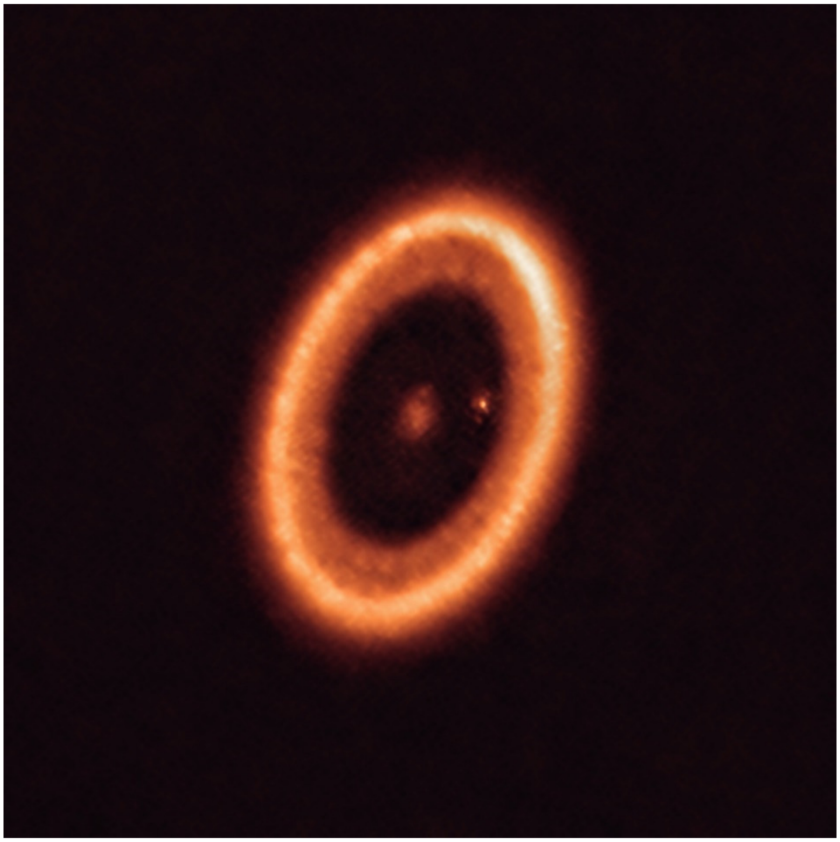


FIGURE 2.7 ALMA image of the young planet-forming star PDS70 showing a outer ring of leftover debris, an inner disk of planetesimals, and a potentially moon-forming disk orbiting a young Jupiter-like planet (at 3:00). The outer ring is about 75 AU in radius, about twice the size of our own solar system. SOURCE: ALMA (ESO/NAOJ/NRAO)/Benisty et al. See https://www.eso.org/public/images/eso2111b/.

2.1.2.2 Atmospheric Spectroscopy to Characterize Worlds

Only a small fraction of the thousands of known planets have been characterized beyond their basic mass or radius, along with orbital properties, but those available observations further illustrate the diversity of worlds. In addition to completing the planetary census, the other major frontier of the 2020s is the spectroscopic observation of planetary atmospheres. A spectrum can yield the abundances of atoms and molecules as well as the temperature of a planetary atmosphere (Figure 2.8). Already several dozen planetary atmospheres have been characterized by the Spitzer and Hubble space telescopes and by large ground-based telescopes. Observations of the transiting and directly imaged planets target the extreme inner and outer reaches of planetary systems, respectively. So far this work has mostly focused on the easier to observe giant planets, but already the diversity of planets beyond what is seen in the solar system is on full display. Detections have been made of clouds of rock dust in the hottest planets, metal vapor such as sodium, potassium, iron, and calcium, escaping hydrogen, and molecules including water vapor, carbon monoxide, and methane. Abundance determinations show the current state of these atmospheres, and—as in the solar system—the history of planetary formation and evolution is embedded in them.

NASA's TESS mission, along with ground-based surveys, are finding planets around the nearby bright stars, to further fuel this revolution in atmospheric spectroscopy. The premier tools for obtaining exoplanet spectra in the 2020s will be JWST and extremely large ground-based telescopes. These telescopes will revolutionize the understanding of the composition of exoplanets. Atmospheres are the

window into many physical, chemical, and formation processes, and these platforms will deliver spectra for a continuum of worlds, for Jupiter-size gas giants, to the Neptune-size planets, to the mysterious mini-Neptunes that dominate the exoplanet census, and down to Earth-size rocky worlds. The lessons of atmospheric physics and chemistry learned from one planetary class can readily inform the understanding of other classes. As astronomers will be far outside of the realm of solar system expertise, and while tentative predictions of what to expect have been made, the joy of discovery will be seeing the diversity of these new worlds. In particular for terrestrial planets, significant progress in the 2020s will occur for systems around low-mass M-dwarf stars. Nearly all of this science will focus on planets within a few tenths of an astronomical unit (AU) of their parent stars, orbits tucked in far closer than Mercury is to the Sun. Moving beyond this to characterization of systems that look more like the solar system will require new capabilities. To understand potentially Earth-like planets around stars like the Sun, which is the only example, so far, for life, will drive the field towards even loftier goals beyond the 2020s (Box 2.1). Are such planets habitable? Do they show signs of life?

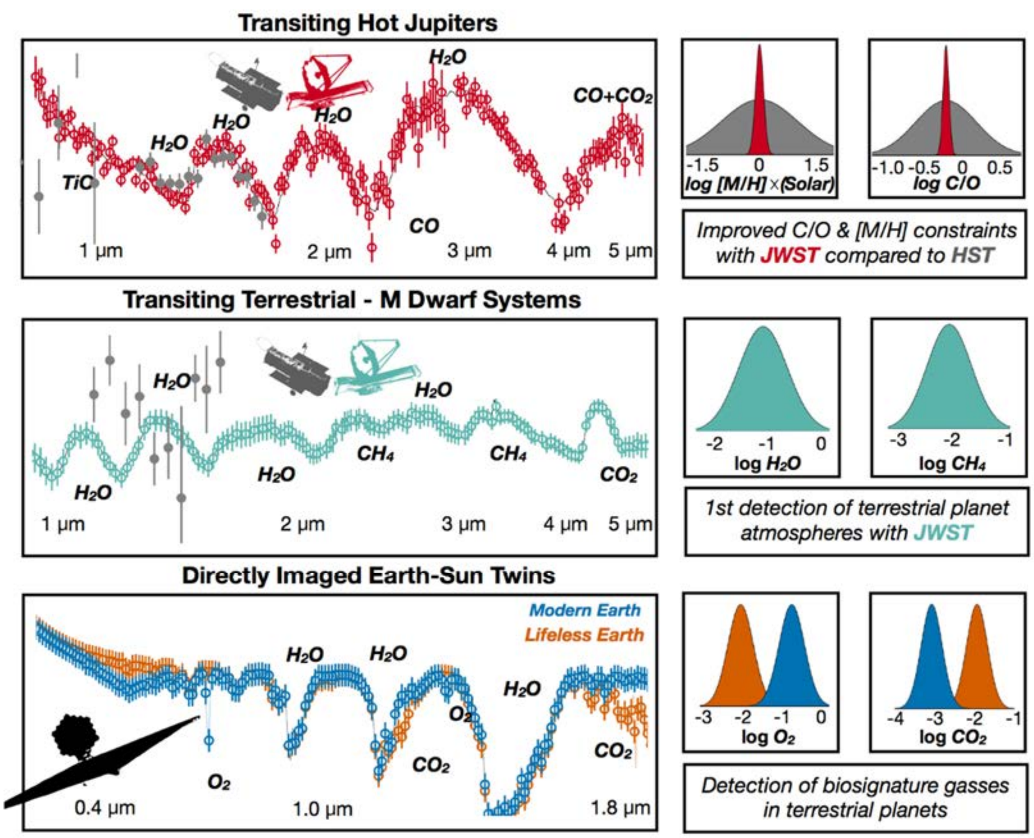


FIGURE 2.8 The 2020s and beyond will be an era of spectroscopy of exoplanet atmospheres. For giant planets such as transiting "hot Jupiters" (*top panel*) the limited wavelength coverage and precision of Hubble will give way to high-precision spectroscopy across JWST's wide wavelength range, yielding the detection of many molecules, and comprehensive atmospheric characterization including metallicities ([M/H]) and carbon-to-oxygen ratios. For terrestrial exoplanets that transit small M dwarf stars (*middle panel*), where Hubble is only able to yield marginal constraints, a significant allotment of JWST observing time will allow for the first reconnaissance of these atmospheres, including the ability to determine the mixing ratios for a range of molecules important for life, like water vapor and methane. Looking to the future (*bottom panel*), to examine the atmospheres of potentially Earth-like planets around Sun-like stars will require further development of a specialized space telescope for high-contrast imaging to measure a reflection spectrum that could show oxygen, methane, water vapor, and carbon dioxide. SOURCE: Courtesy of Natasha Batalha and the PICASO project, http://natashabatalha.github.io/picaso.

BOX 2.1 Detecting Life on Exoplanets

Earth's current atmosphere is shaped in many ways by the presence of life—from the abundance of oxygen, to the atmospheric and climate changes over the past century as humans burn fossil fuels. Such biosignatures could be detected spectroscopically in exoplanets, if life is as prevalent as it is on Earth. Astrobiologists and exoplanetary scientists have assembled lists of proposed biosignatures. Large-scale atmospheric oxygen is one of the most powerful, particularly in combination with the detection of other compounds such as methane. On Earth, atmospheric oxygen would be short-lived if not replenished by photosynthesis. An Earth-like spectrum would be a strong indicator of life-like processes. However, interpretation of such a detection must be extremely careful; abiotic processes can also produce complex chemical signatures. Predicting all the possible chemical pathways in a remote planetary environment, perhaps orbiting a star very different than the Sun, will be challenging. Interpreting any observations as a sign of life must involve high-quality data detecting multiple atmospheric constituents, an integrated view of the planet's evolution and its interaction with its star and solar system, and development of a comprehensive framework allowing a probability analysis. One important component of this will be the study of exoplanets over a wide range of masses, ages, stellar insolations, and compositions; even uninhabitable planets can provide clues to how atmospheres are formed and evolve, as the atmosphere of Venus helped the understanding of the history of Earth.

The search for biosignatures in exoplanet systems is focused on the most Earth-like worlds, particularly those that could have liquid water on their surface. This in turn concentrates attention on planets in the habitable zone. The exact borders of this zone depend on the details of the planet's atmosphere, orbital and rotational motions, and the interactions with the star, but in the solar system at the Sun's current age it extends from just inside the orbit of Earth to somewhere around the orbit of Mars. Since stars with lower mass than the Sun are much less luminous, the habitable zone moves much closer to the star, well inside the equivalent of the orbit of Mercury for the lowest-mass stars.

This range of star/planet separations leads to a range of different pathways for searching for life-bearing planets (Figure 2.1.1). For low-mass stars, the close-in habitable zone means those planets are much more likely to transit their parent star, and the small star yields a larger relative transit signature. These worlds are being identified by ground-based surveys and TESS, and will be studied in transit spectroscopy by JWST and potentially by the ground-based ELTs. Exactly how many Earth-sized habitable zone planets can be probed this way is unclear. It is likely a small number, and not all key biosignatures (including oxygen) may be detectable, but the first glimpses of planets in the habitable zone will come from transiting worlds orbiting the very lowest-mass stars. Directly imaging non-transiting, potentially habitable planets of the nearest low-mass stars requires extreme angular resolution but is only moderately demanding in terms of relative brightness, and hence is feasible from the ground with high-performance adaptive optics. Achieving this is a key science goal of the proposed ground-based extremely large telescopes.

Although the planets orbiting low-mass stars are the easiest to study, those stars are also very different from the Sun, and yield habitable zone environments that may be quite different from the solar system. Low mass stars' high initial luminosities and their prolonged output of high-energy radiation may make atmospheres difficult to keep for habitable zone planets. Potentially Earth-like planets around Sun-like stars orbit at larger distances, and hence are easier to spatially separate from their parent star even with moderately-sized (4 m+) telescopes equipped with coronagraphs or starshades. However, the bright star still makes them hard to see; Earth seen from beyond our solar system is more than 10 billion times fainter than the Sun at visible wavelengths. Reaching this level of planet to star contrast requires extraordinarily precise control of the wavefronts of light and is only practical with dedicated space-based telescopes. At longer wavelengths, the contrast ratio is more favorable, but the thermal background and scattered light would still swamp the planet signal for any but the few very closest sunlike stars.

It will be necessary to study sufficiently large samples of planets both inside and outside the habitable zone to find potentially rare Earth-like planets, so that connections between planetary properties and environment can be explored. Comparative planetology between systems influenced by very different stars and evolutionary pathways and examining multiple planets in a given system will help make the interpretation of atmospheric signatures more certain.

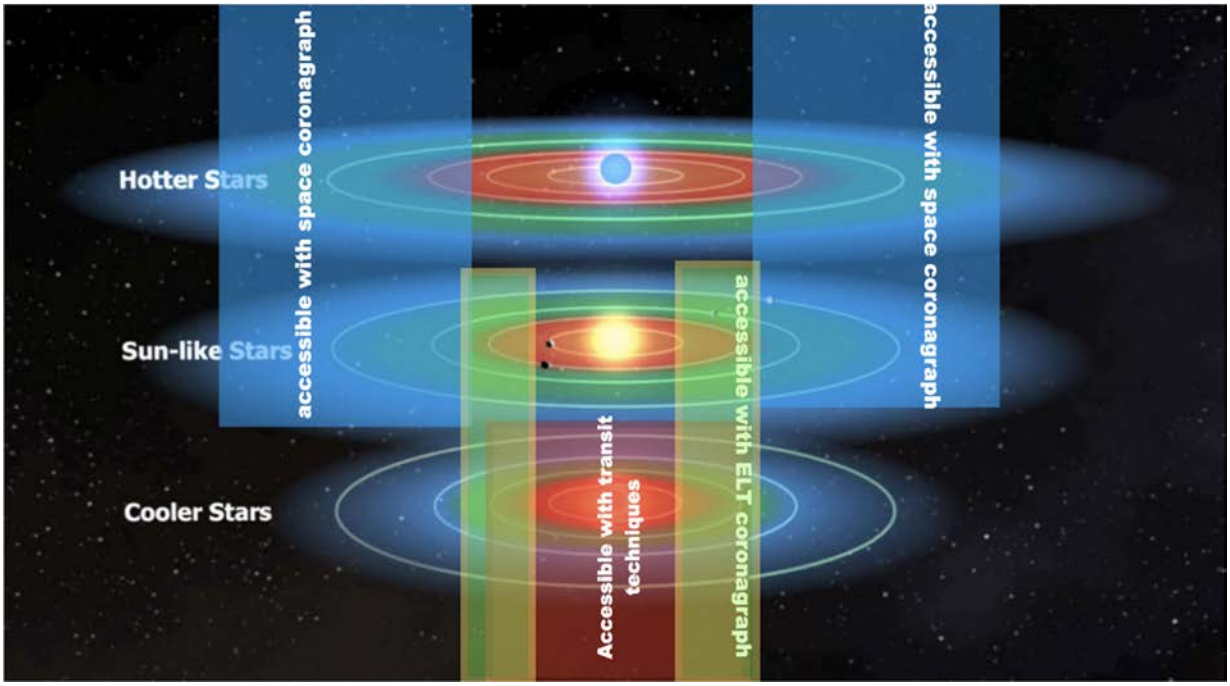


FIGURE 2.1.1 The location around a star where liquid water can exist—the habitable zone—changes with stellar temperature. Different types of telescopes are needed to probe these locations, from space- and ground-based coronagraphs which can return imaging observations of planets, to close-in planets only accessible through transit measurements. SOURCE: NASA/Kepler Mission/Dana Berry, adapted from https://www.nasa.gov/ames/kepler/habitable-zones-of-different-stars.

2.1.2.3 Connections to the Solar System

Studies of exoplanets and of the solar system are tightly intertwined and have enjoyed a profitable give and take in contributing to the understanding of planet formation and atmospheres. Many aspects of exoplanet atmosphere modeling use numerical techniques originally developed for the solar system's own giant and terrestrial planets. The solar system provides "ground truth" to validate spectroscopic observations against in-situ measurements such as elemental abundances on Jupiter or Earth or hazes in satellite atmospheres (Figure 2.9). On the other hand, the vast diversity of exoplanet types—many with no analog in the solar system—and the opportunity to study systems over a range of ages provides insights into formation mechanisms and evolutionary processes that improve the understanding of the solar system's own history.

The solar system contains asteroids and comets as remnants of its original formation. Analogues around other stars take the form of debris disks, massive dust belts produced by collisions among these small bodies. Measurements of the composition, orbital dynamics, and size distributions of small bodies in the solar system provide crucial benchmarks for understanding solar system formation in one spectacularly detailed instance. Time-domain surveys such as the Vera Rubin Observatory's Legacy

Survey of Space and Time will greatly expand the number of known small bodies in the outer solar system, and provide information about its early evolution. Results from recent studies analyzing dynamics of small bodies in the Kuiper Belt provide tantalizing hints, to be confirmed, about the possibility of additional planets.

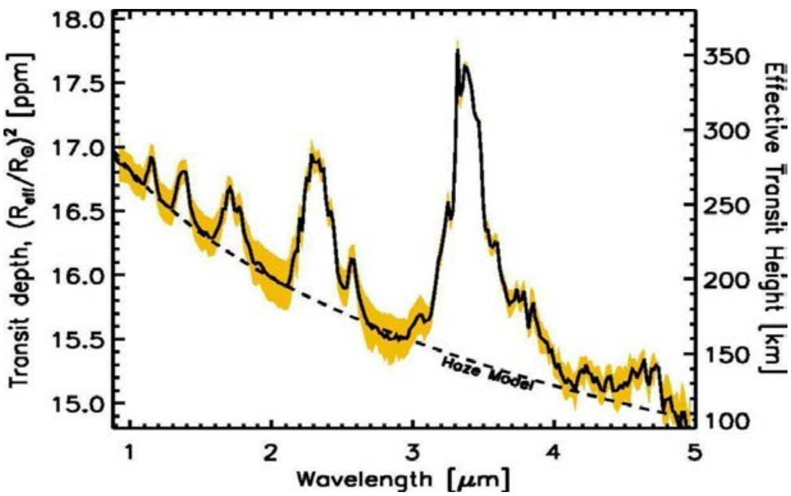


FIGURE 2.9 Spectrum of the hazy satellite Titan during an occultation of the Sun as viewed by the Cassini mission to the Saturn system. Recognizing the analogous geometries between such occultations and exoplanet transits of their host star, Robinson et al. (2014) used these observations to inform how high altitude hazes influence transit spectra of exoplanets. This cross-disciplinary approach will be key to interpreting exoplanet transit spectra taken by the James Webb Space Telescope. SOURCE: Courtesy of T.D. Robinson, L. Maltagliati, M.S. Marley, and J.J. Fortney, *Proceedings of the National Academy of Sciences* 2014, 111 (25) 9042-9047, doi: 10.1073/pnas.1403473111. Copyright 2014 National Academy of Sciences. Reproduced with permission.

While the detailed information available from solar system observations informs theories of planetary system formation in general, such studies are also synergistic. Observations of young planet-forming disks provide a window into the early conditions that led to the formation of the giant planets in this solar system. New generations of radio interferometers will probe inner solar system scales of planet formation, approaching scales of Earth's orbit and super-Earth planet masses. High angular resolution near- and mid-infrared observations of the innermost regions of planet formation around young stars directly image thermal continuum emission of forming planets and provide kinematic constraints. Increased capabilities to track circumplanetary disks and substructure add the possibility to measure orbital motions of Earth-like planets as they are forming, and provide context for factors potentially affecting our early Earth. Detection and study of complex organic molecules provide the chemical initial conditions of forming solar systems. With these high angular resolution and high sensitivity techniques, young planetary systems can be observed—analogs to the solar system—at their moment of formation, validating models that explain the origin of the life-bearing conditions on Earth.

As the only place for which in situ and other detailed studies are possible, advances in knowledge of the solar system are a crucial part of the astrobiological endeavor, and astronomical observations in turn inform the understanding of the planets in direct reach. Future space coronagraphic observations of young Venus or Mars analogs, for example, could help confirm whether those planets had more Earth-like atmospheres in the past. Continued solar system space missions, complemented by long-term monitoring from telescopes at Earth, provide essential details to broadly understand these planets. Comprehensive theories of planet formation must be able to explain both the solar system and the diverse array of known exoplanetary systems.

One of the most exciting astronomical discoveries of the past few years are two interstellar interlopers—the asteroid ‘Oumuamua and the comet 2I/Borisov—that originated around another star and passed through our solar system. Large-scale surveys such as the Rubin Observatory Legacy Survey of Space and Time will discover many more such objects and will provide an increased understanding of context and impact for these interstellar visitors.

2.1.2.4 The Star-Planet Connection

Stars and the planets that orbit them are inexorably connected to each other by their evolutionary history. The star’s properties and evolution influence the evolution and habitability of the planets, particularly of terrestrial planets. The end stages of star formation provide the initial conditions for planet formation. The coming decade will see the implementation of a systems-level approach to understanding the many factors that influence a planet’s habitability.

Most workhorse planet detection methods are a relative measurement, made with reference to the properties of the host star. Stellar surface inhomogeneities pollute the planet measurements due to the combined light of the system and the breakdown of simplistic assumptions about the stellar surface characteristics. Starspots and other variations produce spurious Doppler shifts that can mask Earth-sized planets. Precision planet characterization in the coming decade will motivate the need for better knowledge about the star, particularly its magnetic properties. Conversely, close-in planets can be used as test particles to reveal small-scale stellar inhomogeneities (Figure 2.10), and increase understanding of stars.

A star can influence its near environment by a combination of its radiation, gravity, and particles. The difference between “super-Earth” and “mini-Neptune” planets can potentially be explained by processes driven by the star. A star’s magnetic field is responsible for the heating producing stellar emissions above the photosphere, which manifests as enhanced ultraviolet through X-ray radiation. High energy stellar radiation is a risk to, and potentially a catalyst for, life. The UV emission of a star influences the planetary atmosphere photochemistry, and can create false positives for biosignatures. The extent and amount of high energy radiation from the star determines how much of the atmosphere a close-in planet keeps over evolutionary time. For planet-hosting stars which are cooler than the Sun in temperature, the magnetic field is also complicit in the star losing mass, through a steady stellar wind and potentially transient coronal mass ejections. Eruptive events, characterized as some combination of stellar flaring, coronal mass ejections, and energetic particle events, produce space weather and in the extreme events, influence habitability. Magnetic fields of low-mass stars may prevent some eruptive events from ejecting mass, lending complexity to a blind application of the solar analog, and there are currently few observational constraints on stellar winds or stellar coronal mass ejections. The next decade will see progress in characterizing the habitability prospects, particularly of M dwarf planets. Due to their proximity to the star, they are the most vulnerable to atmospheric loss, coronal mass ejections, tidal heating, and orbital evolution.

Priority Science Area: Pathways to Habitable Worlds

Among the many exciting discoveries and opportunities ahead within the *Worlds and Suns in Context* theme, one question stands out: “Are there habitable planets harboring life elsewhere in the universe?” The answer to this question, and the questions which immediately follow—“Is the Earth unique?” “Are humans alone?”—would have impacts reaching far beyond astronomy, within science and humankind overall. Advances in the understanding of exoplanets and in astronomical instrumentation now allow the planning of major steps towards identifying and studying candidate habitable planets, and making the first tests for the signatures of habitability and life. Laying out the Pathways to Habitable Worlds is the priority science area for this theme.

Permission Pending

FIGURE 2.10 Surface inhomogeneities on a star can be revealed by departures from the expected depth of transiting planets, as in the case of a dark patch and a different light patch on the star as revealed in this example from the transiting planet WASP-19b. SOURCE: Espinoza et al. (2019).

Meeting this ambitious goal will require progress on a variety of fronts, observationally, theoretically, and in the laboratory. Potentially habitable planets are the exception rather than the rule (roughly 20 out of the ~2,400 planets discovered by Kepler, Sec. 2.3.2), and those close enough to be studied in detail by future facilities is smaller yet, so current and planned exoplanet surveys will play a critical role in enlarging the candidate list. Ongoing efforts to characterize the surfaces and atmospheric compositions of known exoplanets, from space and the ground, will be important for quantifying the demographics of the overall planet population and refining the diagnostic tools which will eventually be brought to bear on the candidate habitable worlds.

Observing and characterizing candidate habitable planets themselves, however, whether by direct imaging or indirect observations in transiting systems, requires new telescopes on the ground and in space. The first target systems are likely to be planets orbiting close to the lowest mass and faintest stars, M-dwarfs. Their low luminosities (thousands of times less than the Sun) make it easier to detect an

orbiting planet against the stellar glare, but the habitable planet zones will lie close to the stars themselves. JWST may be able to measure a few such systems in transit, and spectroscopic observations of larger samples form a keystone part of the science cases for the 24-40m ELTs (E-ELT, GMT, TMT).

A longstanding goal ever since the discovery of the first exoplanets is to image and characterize planets orbiting Sun-like stars, including those in the habitable zone. Such stars have long been suspected as offering the most likely sites not only for liquid water but also life, and space-based telescopes with sufficient stability and high-contrast capabilities to image and spectroscopically measure planets around the nearest such stars are now within our reach. A space telescope similar in wavelength coverage to the Hubble Space Telescope, and with an aperture of at least 6m and coronagraphic imaging capability should be capable of observing approximately 100 nearby stars, and successfully detect potentially habitable planets around at least a quarter of the systems. Such an observatory would also provide valuable information on other extrasolar planets, and be versatile enough to carry out groundbreaking observations of stars, galaxies, black holes, and the gases and baryons within and between galaxies, with a scientific impact rivaling that of previous “great observatories” such as HST.

The potentially habitable worlds ripe for discovery in this decade and the next represent only the tip of a deep iceberg in exoplanetary exploration. Studies of Super-Earths and mini-Neptunes as well as larger Jupiter-sized planets fill in critical areas; these measurements can be done with a range of facility sizes and round out a comprehensive approach to the subject. Considering solar systems as a whole and understanding how the individual components interact with each other is a crucial component of understanding the processes at work to arrive at the observed state. While the ultimate quest is to find potentially habitable planets, the perspective of a wide variety of planetary demographics, characteristics, and evolutionary paths (including those that lead away from habitability) is needed to understand where the multiple highway exits lead along the route to answering the question “Are humans alone?”

The path to habitability starts at the beginning of a planet’s journey and requires investigation of the chemical and dynamical processes at work to determine conditions early on in a planet’s life that lead both towards and away from habitability. Improvements in imaging and spectroscopy with future large facilities at mm and sub-mm wavelengths will probe the rings and gaps caused by forming planets in the disk of gas and dust around a young star, and create a census of the properties of these forming planets. Spectroscopic probes of planet-forming gas reveal the chemical initial conditions that solar systems and individual planets in formation experience; the study of complex prebiotic species paints the picture of chemical evolution pathways needed at the beginning of this path. Measuring the water reservoirs in star- and planet-forming disks is crucial for understanding the mechanisms by which terrestrial planets gain their water.

Although this mapping of the Pathways to Habitable Worlds has emphasized the central roles to be played by major existing and planned observatories, success will also require a wide array of enabling and foundational projects and studies. These include observational and theoretical studies of the linking between star and planet formation, the chemical processes—both inorganic and organic—critical for planet formation, the evolution of planetary atmospheres, and the origins of life, laboratory measurements of spectroscopic tracers and astrochemical processes, and detailed study of stellar activities, variability, and surface structures, all of which can mimic observational signatures of transiting

2.2 NEW MESSENGERS AND NEW PHYSICS

Our understanding of the universe has been repeatedly transformed by observing across the entirety of the electromagnetic spectrum, from the radio to the gamma-rays. These observations can test or reveal physics in ways that are not possible on Earth. Now, nearly daily movies of the sky, as well as the new messengers of neutrinos, particles, and gravitational waves, provide new ways of doing astronomy and new methods for uncovering and testing new physics.

Observing the universe in new ways has historically involved expanding our observations to cover the full electromagnetic spectrum, not just the part accessible with our eyes. In doing so, X-ray observations revealed the dynamic coronae of the Sun and other stars, accreting neutron stars and black holes, and the hot plasma that pervades interstellar and circumgalactic environments. Radio observations revealed the existence of neutron stars, whose remarkably stable rotation rates have since been used to discover planets and confirm the theory of general relativity's prediction of orbital decay via the emission of gravitational waves. And telescopes observing in the infrared can peer into the enshrouded stellar nurseries where stars and planets form.

New views of the universe also come from observing in new ways—for example, by making observations with much higher sensitivity, angular resolution, or time resolution, or by obtaining higher dynamic range views of previously hidden phenomena. The movies of stars orbiting the 4 million solar mass black hole in the center of the Milky Way Galaxy would not have been possible without adaptive optics and near-infrared interferometry to overcome the blurring effects of Earth's atmosphere. More recently, the Event Horizon Telescope's unprecedented angular resolution at millimeter wavelengths (via interferometry) provided the first direct image of the near-horizon environment of a black hole (see Figure 2.11), captivating scientists and the public alike.

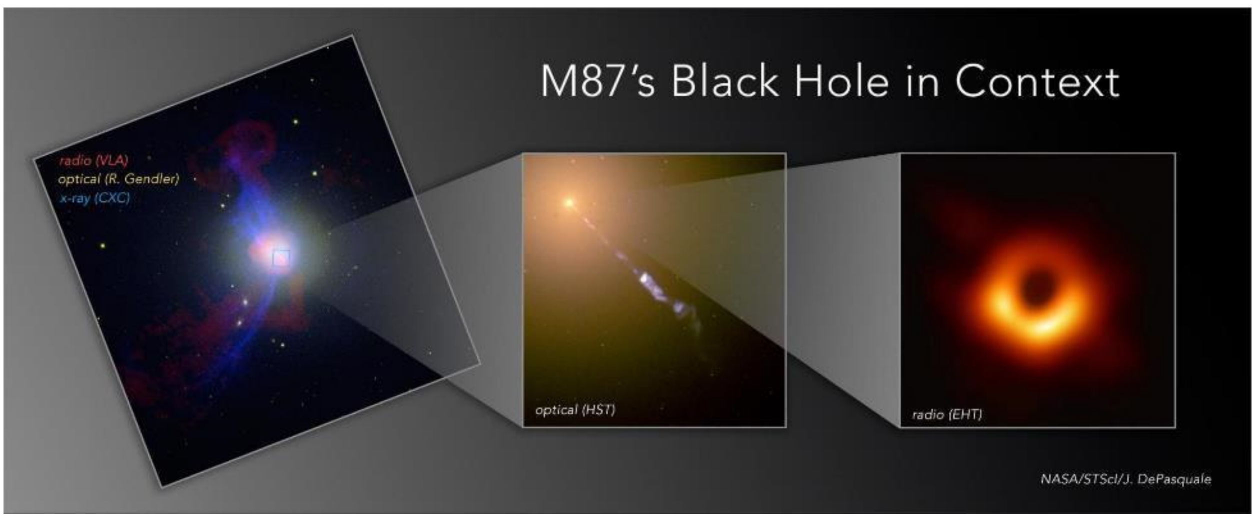


FIGURE 2.11 The galaxy M87 (*left*) harbors a 6.5 billion solar mass black hole at its center that produces a spectacular jet (*middle*). The unprecedented angular resolution at millimeter wavelengths of the Event Horizon Telescope produced the image of electromagnetic radiation from plasma near the horizon of the black hole (*right*). The image shape is interpreted as being due to motion of the radiating plasma at near the speed of light (producing the brighter lower half) and strong gravitational lensing by the black hole (producing the 'shadow of the black hole,' the deficit of light at the center). SOURCE: *Left:* Adapted from Chandra X-Ray Observatory, "M87: A Nearby Galaxy Metropolis," https://chandra.harvard.edu/photo/2008/m87; X-ray: NASA/CXC/CfA/W. Forman et al.; Radio: NRAO/AUI/NSF/W. Cotton; Optical: NASA/ESA/Hubble Heritage Team (STScI/AURA), and R. Gendler; Middle: Adapted from Hubblesite, "Black Hole-Powered Jet of Electrons and Sub-Atomic Particles Streams from Center of Galaxy M87," https://hubblesite.org/contents/media/images/2000/20/968-Image.html, NASA and The Hubble Heritage Team (STScI/AURA). Right: Adapted from The Event Horizon Telescope Collaboration, "First-ever Image of a Black Hole Published by the Event Horizon Telescope Collaboration," https://eventhorizontelescope.org/blog/first-ever-image-black-hole-published-event-horizon-telescope-collaboration.

By observing ever fainter sources across the electromagnetic spectrum, it is also possible to peer back into the distant past when the universe and its constituents were young, and unravel the history of the universe. Perhaps most spectacularly, observations of the cosmic microwave background radiation

measured the geometry and mass-energy content of the universe and helped transform cosmology into a precision science.

The most radically different views of the universe are provided by messengers other than electromagnetic radiation in the form of photons or waves. Neutrino observations confirmed the theoretical prediction that hydrogen fusion powers the luminosity of the Sun, and demonstrated that neutrinos have mass, a key insight into physics beyond the standard model of particle physics. Cosmic-ray measurements have found particles whose energies are enormous compared to those that can be produced in the Large Hadron Collider (LHC) at CERN, but whose astrophysical origin remains a mystery. In 2013, the south pole IceCube observatory detected a diffuse high energy neutrino flux of astrophysical, but unknown, origin. Starting in 2015, LIGO opened up the gravitational wave view of the universe by detecting merging binary black holes. The simultaneous detection of gravitational waves and electromagnetic radiation from a binary neutron star merger in 2017 showed the power and complementarity of multi-messenger observations (see Box 2.2).

As views of the universe have expanded so has astronomy's impact on basic physics. Many frontier science questions identified by the Survey's science panels center on the intertwined themes of using new techniques to see the universe in new ways (new messengers) and uncovering new physics with advanced astronomical observations. In what follows the presentation of this science theme is organized by first discussing cosmology and the dark sector and then turning to the new astronomy enabled by observations with particles, neutrinos, gravitational waves, and light.

2.2.1 Cosmology and the Dark Sector

The fundamental paradigm of modern cosmology is the Hot Big Bang, in which an initially hot, dense, and nearly smooth universe rapidly expands and cools. All evidence suggests an initially simple universe, made of a nearly uniform collection of light nuclei and electrons, a sea of radiation, a similar sea of cosmic neutrinos, and dark matter of an unknown nature. As time passed, small initial differences in density grew under the action of gravity to form the rich structure described in the *Cosmic Ecosystems* science theme of this report. The transition from a smooth universe to one with stars and galaxies occurred less than 500 million years after the Big Bang. Finding innovative ways to probe cosmology in the "dark ages" prior to any significant star formation is one of the discovery areas identified here. The development of "LambdaCDM," our current standard cosmological model is one of the major intellectual triumphs of the past few years; the nature and origin of its key ingredients—dark matter, dark energy, and a nearly scale-invariant spectrum of primeval mass fluctuations—remain, however, some of the biggest mysteries in science.

Observations of the motions within galaxies and clusters to the large-scale distribution of intergalactic gas and the CMB require more matter than can be explained by the observed baryons. In the common interpretation, this is cold dark matter, an unseen gravitating material—nearly 10 times more abundant than baryonic matter—that moves non-relativistically in the recent universe. Over the many decades since dark matter was posited by Fritz Zwicky to explain galaxy motions in the Coma cluster and by Vera Rubin to understand galaxy rotation, astronomers have learned primarily what it is not: not like anything that has been seen on Earth. It could be a particle with the mass of a proton that does not significantly interact with normal matter or, at another extreme, it could be a "particle" with a quantum-mechanical wavelength the size of a small galaxy. While physicists attempt to identify dark matter through ambitious and excruciatingly careful laboratory experiments, astronomers in the coming decade will wield the threefold tools of theory, simulation, and observations to search in parallel. Dark matter could leave detectable traces of its potentially more complex interactions through deviations from the simplest version of the cold dark matter paradigm or through emission of unexpected particles (gamma-rays, positrons, narrow radio frequency lines) produced by dark matter interactions. The signatures of complexity in the properties of dark matter could come from the most distant sources (e.g., the CMB) or some of the nearest (e.g., nearby wide stellar binaries or the internal kinematics of dwarf galaxies). Nearly

all astronomical facilities—current, imminent, and aspirational—contribute to the study of dark matter. New large ground-based optical-infrared telescopes would be particularly impactful, e.g., by studying the internal motions of stars in dwarf galaxies and testing the nature of the dark matter that holds those galaxies together.

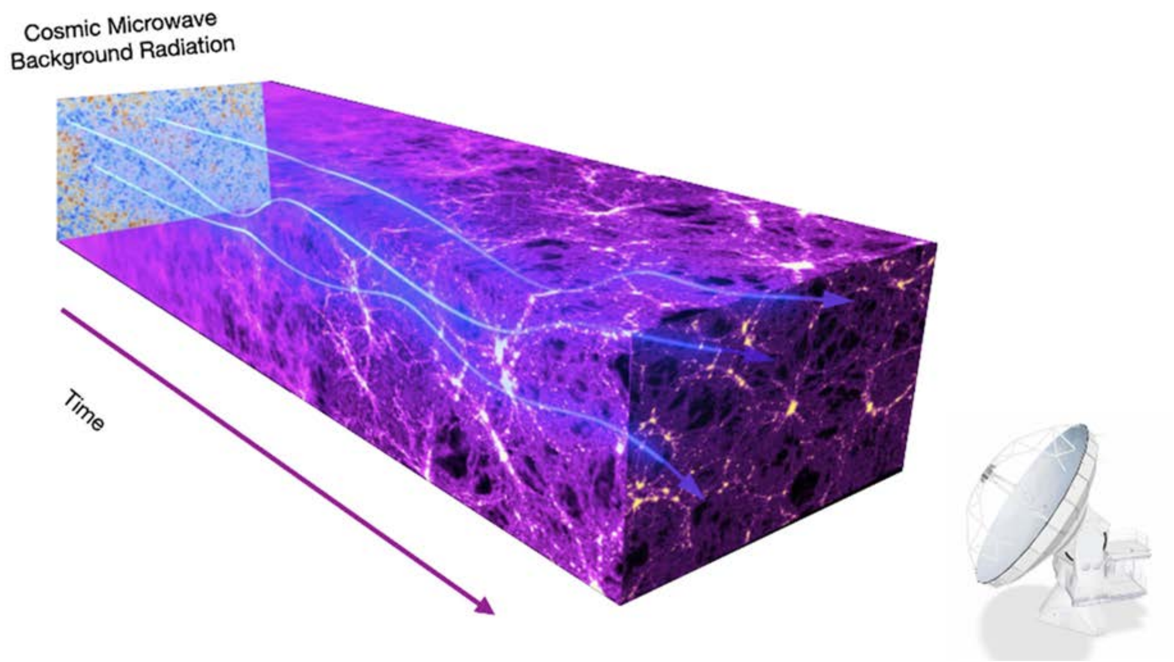


FIGURE 2.12 The cosmic microwave background (CMB) can be thought of as a backlight at the edge of the observable universe as depicted above. Light from it traverses the universe on its way to us on the right. Concentrations of mass, shown here as the brighter areas in the web of dark matter, deflect the light through the process of gravitational lensing. The image of the CMB that finally reaches Earth is then a distorted view of the true CMB, like the view through an imperfect piece of glass. Measuring these distortions determines the three-dimensional mass distribution that the light has passed through, which is shaped by the distribution of dark matter throughout the universe. The light of distant galaxies can be distorted in a similar manner. SOURCE: *Left:* Copyright ESA and Planck Collaboration. *Right:* ALMA antenna, https://public.nrao.edu/telescopes/alma/, National Radio Astronomy Observatory, Associated Universities, Inc., and the National Science Foundation.

Even with the current poor understanding of what dark matter *is*, there is broad consensus around how it behaves gravitationally on large scales. This allows theoretical models to robustly connect the final distribution of dark matter to the conditions in the very early universe. Astronomers are on the brink of using this connection to make 3D maps of the dark matter distribution and use those maps as probes of fundamental physics. The properties of the dark matter distribution can be traced by the ways it subtly shifts light rays as they propagate through the universe, through the effect of gravitational lensing (Figure 2.12). Analyzing the fine details of maps of the CMB or of the shapes of distant galaxies, reveals the 2D matter distribution, which can then be deprojected into the full 3-dimensional view. This is one of the primary goals of Roman, Rubin, Euclid, and new more powerful CMB instruments. The statistical properties of these maps will eventually be able to measure the sum of the masses of the neutrinos, which will be yet another major contribution of astronomical observations to basic physics. In parallel with lensing studies on large scales, more detailed, focused studies of individual lensed galaxies, supernovae, and quasars (accreting black holes), and even lensed stars at cosmological distances, will test the predictions of the cold dark matter paradigm on the smallest scales, where we expect deviations to be most evident.

Along with dark matter, the second great cosmological unknown is the nature of the “dark energy.” Observations of the recent universe have shown that its expansion is accelerating. This remarkable discovery is not explained by a model containing only matter (even dark matter), but instead indicates a new feature, dubbed dark energy. Arguably the simplest explanation for dark energy is Einstein’s cosmological constant—an energy and pressure characterizing empty space whose gravitational effect drives the acceleration—but the incredibly small value, compared to what is expected from quantum theories, leads physicists to think that dark energy may be more complicated. The cosmological constant predicts a specific form for the acceleration in the expansion of the universe as a function of time. Observations in the coming decade using Rubin, Roman, Euclid and higher resolution, higher sensitivity CMB facilities will test whether observations are, or are not, consistent with a cosmological constant. These results will be a major legacy of the *New Worlds, New Horizons* decadal survey. Any deviations from a cosmological constant model could touch off a second revolution as powerful as the initial discovery of the accelerating universe itself.

While the next generation tests of dark energy are underway, astronomers are already finding hints that the current cosmological model may be incomplete. An unexpected tension has developed between two different ways of determining the present expansion rate of the universe. In one, based on measuring distances and velocities in the local universe, the current expansion rate is roughly 5 percent higher than inferred from the second method, which uses the standard cosmological model to extrapolate from early universe CMB measurements to today. Continued measurements from satellites and the ground, including those using gravitational waves, will be able to distinguish between a systematic effect in one of the methods or the need for a previously unknown component of the universe.

The fluctuations seen in the CMB are believed to have been imprinted in the earliest phases of the Big Bang during a period of cosmological inflation in which extraordinarily rapid expansion established the large-scale homogeneity and flatness of the universe while also causing quantum fluctuations which subsequently grew into the fluctuations we observe. One of the most exciting opportunities in the coming decade is that CMB measurements may reveal remnant gravitational waves from this early epoch, as depicted in Figure 2.16 below. The presence of gravitational waves would manifest as a distinctive polarization pattern in the CMB, called “primordial B-modes,” at angular scales of 3 degrees and larger. These scales are accessible from the ground and space. If primordial B-modes are found they will provide critical constraints on the physics of inflation and give new insights into physical processes at energy scales orders of magnitude larger that can be attained at the LHC. Efforts are already underway to detect them, but higher angular resolution and sensitivity CMB observations from the ground and in space will be needed given the difficulty of detecting the small B-mode signal amidst the polarized galactic foreground.

In summary, the standard cosmological model is both a remarkable triumph and an astonishing puzzle. With a relatively modest number of parameters, it continues to match observational results despite orders of magnitude improvement in cosmological measurements over the past twenty years. However, there is a mystery novel’s worth of hints that the same model is incomplete, as the most important components are not yet understood. This represents one of the great contributions of astronomical observations to basic physics, and a continued opportunity for breakthroughs and surprises in the coming decades. Unraveling the many cosmological mysteries will require a particularly close interplay between theory, simulation, observations, and laboratory experiments.

2.2.2 New Messenger

Astronomy has long been a science rooted in the observation of light (photons). The past decade, however, has overturned this understanding of what astronomy is and could be, thanks to observations with new messengers that carry new information about the workings of the universe. Gravitational waves, neutrinos, and cosmic rays (Figure 2.13)—long viewed as largely the province of physics—have all now passed into the realm of astronomy. This is due to multiple breakthrough discoveries in the last decade,

using facilities such as Auger, IceCube, and LIGO. At the same time, astronomy's traditional pursuit of photons is being transformed by new observational facilities that probe time variable and transient phenomena. Characterizing the time-variable electromagnetic universe has become increasingly sophisticated, thanks to pathfinding optical telescopes such as the All-Sky Automated Survey for Supernovae (ASAS-SN) and the Zwicky Transient Facility (ZTF) that are setting the stage for the Rubin in the coming decade. Outside the optical, dedicated radio surveys with the Karl Jansky Very Large Array (JVLA) and the Canadian Hydrogen Intensity Mapping Experiment (CHIME), and other international facilities are uncovering new and unexpected phenomena, such as fast radio bursts, while high energy space telescopes sensitive to explosive events like gamma ray bursts become ever more central to interpreting signals from new cosmic messengers. It is not an exaggeration to say that nearly daily movies of the sky made across the electromagnetic spectrum are their own form of "new messenger," with information that is fundamentally distinct from a static view of the universe. This led to the identification of time domain astronomy as a key discovery area in *New Worlds, New Horizons*.

These seemingly separate advances in observational techniques are in fact intimately related: most of the known and anticipated sources of gravitational waves, neutrinos, and cosmic rays are also time variable or transient electromagnetic sources (e.g., neutron star mergers, gamma-ray bursts, black hole jets, and stellar explosions). Combining information from all messengers can unravel the physics at the heart of these objects, as was demonstrated so spectacularly in the case of the binary neutron star merger GW170817 (Box 2.2).

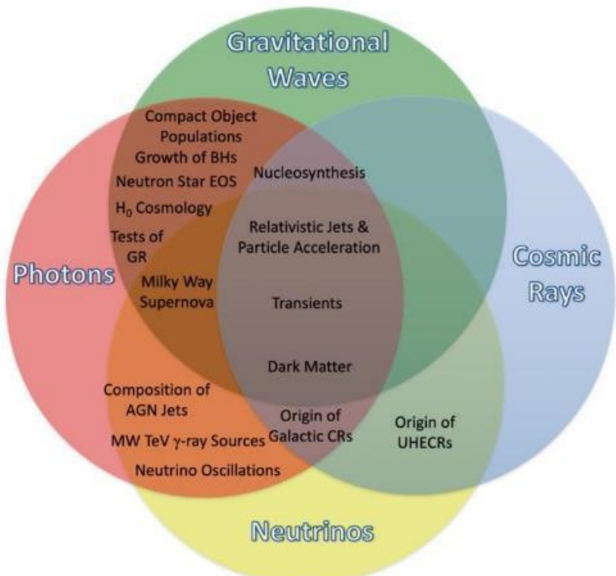


FIGURE 2.13 The combination of multiple messengers provides unique insights into astrophysical sources, particularly those involving strong gravity or relativistic motion (see Box 2.2). SOURCE: Report of the Panel on Particle Astrophysics and Gravitation.

BOX 2.2 Multi-Messenger Astronomy

The binary neutron star merger GW170817, detected both in electromagnetic and gravitational waves, was a watershed event that confirmed the long-anticipated promise of multi-messenger astronomy. The confluence of decades of work in theory, numerical relativity, nuclear astrophysics, gravitational wave detectors and analysis methods, combined with measurements across the electromagnetic spectrum from space and ground, building on the international network for rapid follow up originally developed for the study of gamma ray bursts, produced what has become the archetype for multi-messenger astronomy, a field destined to blossom in the coming decade.

In GW170817, gravitational wave measurements determined the mass of the merging neutron stars and an initial sky localization, while electromagnetic observations determined the host galaxy of the merger and the mass, speed, energy, and composition of matter ejected from the system during the merger (Figure 2.2.1). This ejecta consisted of both a jet of relativistic material that powered non-thermal radiation from the radio to the gamma rays and slower more spherical ejecta that powered the thermal optical and infrared light. The data indicated that the latter was a “kilonova,” optical/infrared light powered by nuclear decays involved in the production of many of the heaviest elements in nature. Indeed, the optical-infrared light provided strong evidence that neutron-star mergers are a significant astrophysical site for the production of rapid neutron capture elements (including the rare Earth metals, platinum, and gold), a long-standing mystery in our understanding of the origin of the elements traced in the spectra of stars.

Permission Pending

FIGURE 2.2.1 Schematic of the binary neutron star merger GW170817 observed in gravitational waves and light (radio to gamma-rays). SOURCE: Margutti and Chornock (2021), ARAA.

The combination of a gravitational wave distance to the merger and a redshift in the spectrum of the host galaxy also allowed a fully independent measurement of the Hubble constant, the value of which is a source of uncomfortable cosmological tension and in need of new measurements (See Section 2.2.1). Although the single measurement with GW170817 is not as precise as other techniques, multi-messenger cosmology will increase in importance in the coming decade as we detect ever more binary neutron star and black hole mergers.

Observations of SN 1987A and the Sun in light and neutrinos and GW170817 in light and gravitational waves revealed the power of multi-messenger astronomy. This is but a taste of the feast that is to come. Higher sensitivity high-energy neutrino experiments will detect individual astrophysical sources. A supernova in the Milky Way Galaxy would be a multi-messenger, multi-wavelength goldmine, detectable in light, neutrinos and possibly gravitational waves. A next generation ground-based gravitational wave network could detect and localize every solar-mass binary black hole merger in the universe, transforming astronomy and cosmology. Pulsar timing and the space-based interferometer LISA will open up other parts of the gravitational wave spectrum, revealing new sources and new surprises, much as the first X-ray and radio telescopes did. Capitalizing on these opportunities will also require a new generation of theoretical and computational models that combine general relativity, nuclear astrophysics and plasma physics. Likewise, new electromagnetic facilities are needed, including those with transient capabilities, larger ground-based optical-infrared telescopes for detailed spectroscopic follow-up, new cm-wavelength radio arrays for observing the non-thermal radiation from jets, and new space-based satellites to provide critical pieces of the picture missing from the ground (e.g., the ultraviolet and gamma rays).

2.2.3 Neutrinos and Cosmic Rays

Trillions of neutrinos stream through us from the Sun every second, with a similar number from the cosmic neutrino background. This is an indication of just how difficult neutrinos are to detect relative to how common they are, but in this difficulty lies their promise. Precisely because neutrinos interact so little with matter, neutrinos carry information about the inner workings of some of nature's most important energy sources: the nuclear furnaces inside stars, the formation of neutron stars in stellar explosions, and the conditions in the jets of relativistic particles that originate near the event horizons of black holes in galaxy nuclei. Neutrinos that are detected also point directly to the celestial position of their source. In the next few decades, astronomical observations will begin to routinely measure the cosmos with these most elusive of particles. Here it is important to distinguish between cosmological neutrinos (which, like the CMB, have redshifted to millielectronvolts), neutrinos emitted by fusion processes in stars, which are in the megaelectronvolt range, and neutrinos in the gigaelectronvolt range and above, which are produced by hadronic collisions between cosmic rays and ambient matter. We focus on the lattermost here, as they are an unambiguous signature of ion acceleration, follow straight line orbits from their source to our detectors (unlike cosmic rays themselves, which are scrambled by magnetic fields), and unlike their photon counterparts, are immune from optical depth effects.

Theoretical models of the early phases of the Big Bang predict that there should be a nearly uniform cosmic neutrino background similar to the CMB. There are laboratory experiments aimed at detecting the cosmic neutrino background but it is an extremely challenging measurement. In contrast to the sea of low energy neutrinos produced in the Big Bang, higher energy neutrinos are messengers from some of Nature's most dramatic events. When the relatively nearby supernova 1987A exploded, some 25 neutrinos were measured in the Kamiokande, IMB, and Baksan neutrino detectors, which were state of the art for their times. Thirty-five years later, modern neutrino detectors would measure tens of thousands of neutrinos from a supernova in the Milky Way. As similar supernovae explode throughout the universe, they collectively produce a diffuse background of megaelectronvolt (MeV) neutrinos, one that should be within reach of forthcoming experiments if current theories are correct. Detection of either a galactic

supernova or the diffuse background would test models of the formation of neutron stars and black holes and the core-collapse explosion mechanism. Only neutrinos and gravitational waves can peer into the inner regions of these events where densities reach nuclear scales.

The skies are full of even more energetic sources that likely produce neutrinos, some continuous, some episodic, and some transient. For example, relativistic jets—collimated beams of ejected material moving at nearly light-speed—are known to emanate from supermassive spinning black holes in active galactic nuclei (see Figure 2.11). These jets span up to millions of light years in extent. Similar relativistic jets are also associated with gamma ray bursts (GRBs), extremely energetic events that emit in just seconds the same amount of energy the Sun will emit over its lifetime. This rich population is likely to be a significant source of neutrinos detectable with future facilities.

The prospects for future neutrino astrophysics are promising. Over this past decade, the IceCube experiment at the South Pole detected an unresolved extragalactic background of 60 neutrinos with energies in the teraelectronvolt to petaelectronvolt (TeV-PeV)[2] range. Their distribution on the sky indicates that they are produced by distant sources well outside our galaxy, and thus are likely to be by products of energetic events throughout the universe (much like diffuse X-ray and gamma ray backgrounds). To date, only the Sun (Figure 2.14), SN1987A, and potentially one blazar have been imaged in neutrinos of any energy, but with new facilities the excitement of identifying specific individual sources is likely just beginning.

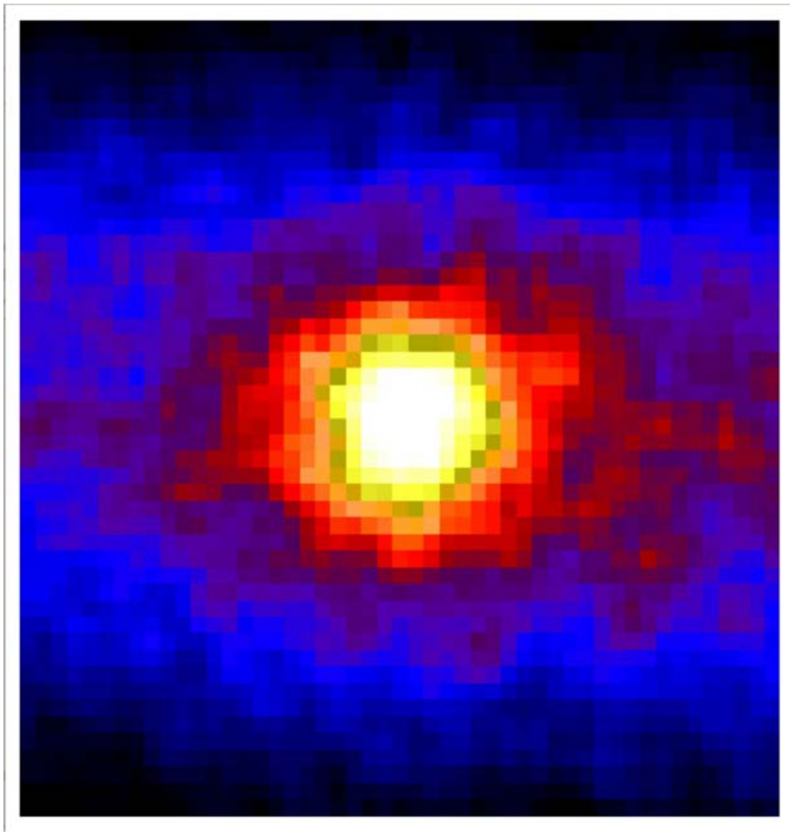


FIGURE 2.14 An image of the Sun taken with the Super-Kamiokande neutrino detector in Japan. These data were collected from neutrinos emitted from the core of the Sun, which after traversing the Earth-Sun separation, travelled through Earth to reach the detector. SOURCE: http://strangepaths.com/the-sun-seen-through-the-earth-in-neutrino-light/2007/01/06/en/, Kamioka Observatory, ICRR (Institute for Cosmic Ray Research),The University of Tokyo.

[2] A teraelectronvolt (TeV) is 10^{12} eV, a petaelectronvolt (PeV) is 10^{15} eV, and an exaelectronvolt (EeV) is 10^{18} eV. The protons in the Large Hadron Collider in CERN have an energy of about 10 TeV, orders of magnitude below those of the highest energy cosmic rays.

Neutrinos are not the only messengers of relativistic and energetic astrophysical phenomena. Quite independently, other experiments have detected ultra-high-energy cosmic rays (UHECRs, in the ~10^{18} eV [EeV] range). What produces these remarkably energetic particles? Are the TeV-PeV neutrinos and ultra-high-energy cosmic rays produced in the same sources? They are widely surmised to be accelerated in the relativistic jets of accreting supermassive black holes or gamma-rays bursts, but this has yet to be tested observationally. Unfortunately, direct identification of a cosmic-ray source is difficult, since UHECRs are charged particles and are thus deflected as they travel through magnetic fields that permeate the universe. However, a clear directional signature would be provided by the high-energy neutrinos that the UHECRs produce in the regions where they are accelerated. Higher sensitivity neutrino observations with better sky localization are critical for unraveling how nature's most extreme particle accelerators work.

2.2.4 Gravitational Waves

Gravitational waves are the newest detectable messenger traveling through the astronomical landscape. They provide a unique probe of regions with large amounts of mass moving at near the speed of light: black hole and neutron star collisions, neutron star and black hole formation in stellar explosions, and the first fractions of a second of the Big Bang (Figure 2.15). The importance of gravitational waves lies in part with the central role that black holes and neutron stars play in many areas of astronomy, from stellar evolution to galaxy formation. In just the five years since LIGO's first detections were announced, gravitational wave measurements have already left astronomers in awe, with insights into the origin of neutron-rich elements (Box 2.2), the detection of stellar-mass black holes much more massive than previously known, and new tests of gravity in the strong field regime.

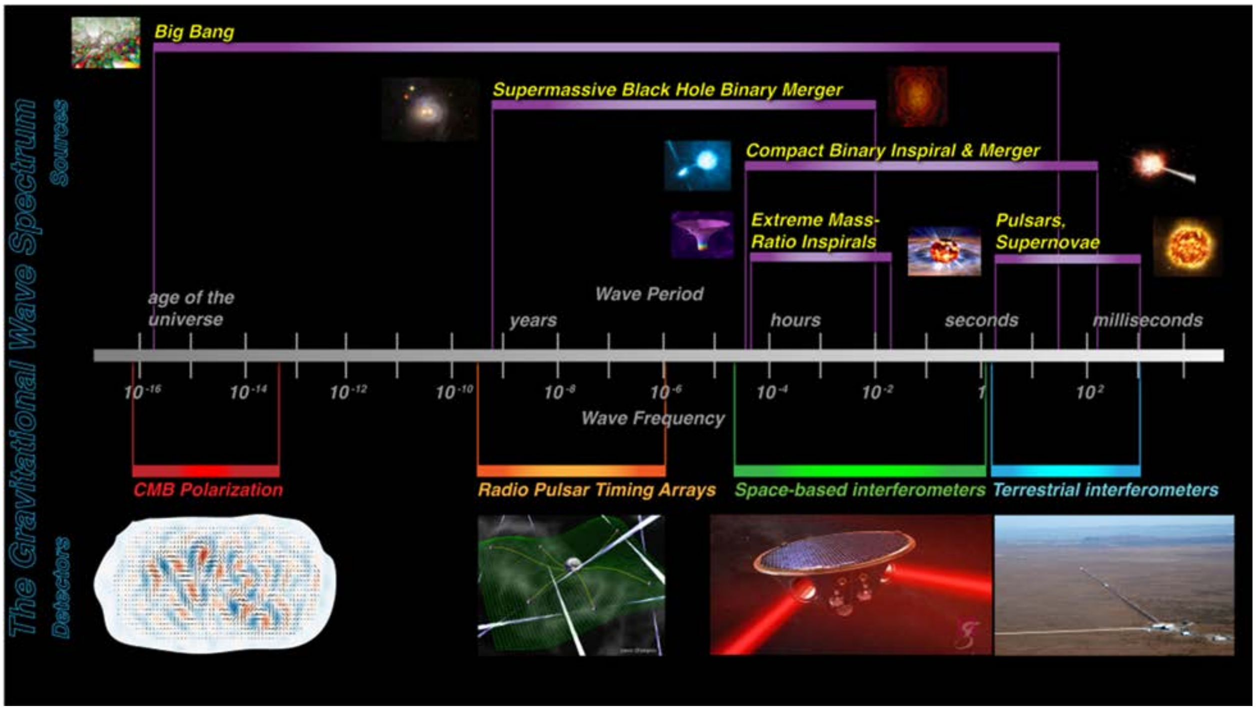


FIGURE 2.15 The full spectrum of gravitational-wave emission. The diversity of objects and events at the top of the image produce gravitational waves of different frequencies that need different detection instruments shown at the bottom. SOURCE: NASA/J. I. Thorpe.

Similar to the electromagnetic spectrum, a variety of astronomical phenomena produce a broad frequency range of gravitational waves (see Figure 2.15). The kilometers-long ground-based detectors LIGO and Virgo are sensitive to signals from relatively small systems: neutron stars and stellar mass black holes with radii from 10 to a few hundred kilometers, and with masses up to several hundreds of solar masses. Colliding massive black holes with millions of solar masses at the centers of galaxies and inspiraling white dwarfs are thousands to millions of kilometers in extent, and need space-based detectors, such as the LISA mission led by ESA with NASA contributions, scheduled to launch in the mid 2030's. To detect the gravitational waves from yet larger objects, like black holes weighing billions of solar masses, requires galaxy-sized baselines. This is done with high precision timing of the arrival of radio pulses from pulsars distributed throughout our galaxy. The loss of Arecibo is a setback for pulsar timing experiments, and underscores the need for radio facilities that can continue this critical science.

The larger and more sensitive ground-based gravitational wave interferometer network anticipated in the coming decade will detect a large number of stellar-mass black hole mergers and determine their masses and spins. Careful comparison of the properties of these mergers with theoretical models will inform which arise from binary or triple stellar evolution, which arise in dense stellar systems like globular clusters, and which might have entirely different origins. The population of neutron star and black hole mergers with masses of ~2-5 M_{sun} in the 'mass gap' between neutron stars and black holes will provide key constraints on our understanding of massive stellar evolution, the maximum mass of neutron stars, and core-collapse explosion physics.

Combined gravitational wave and electromagnetic observations have the potential to finally crack the longstanding puzzle of the origin and growth of massive black holes, one that lies at the intersection of the understanding of stars, galaxies, accretion disks, and cosmology. LISA and future ground-based gravitational wave detectors will detect black hole mergers in the earliest phases of galaxy formation less than a billion years after the Big Bang, while LISA, ground-based detectors, and pulsar timing arrays will constrain the merger rate over cosmic time for a range of black hole masses. Combining this gravitational wave data with deeper electromagnetic observations of accretion onto black holes, particularly in the infrared (e.g., JWST) and X-ray, will inform whether massive black holes originate from the remnants of massive stars. Are they built up by stellar and black hole collisions in dense star clusters, or perhaps they formed from the direct collapse of gas clouds in some of the first galaxies? Gravitational wave measurements will also be critical for disentangling the role of mergers and gas accretion in growing black holes over cosmic time.

The simultaneous detection of gravitational waves and light from the binary neutron star merger GW170817 was a transformative event (Box 2.2). More sensitive ground-based gravitational wave interferometers will provide a much larger sample of binary neutron star and neutron star-black hole mergers. The light from GW170817 was powered by material ejected during the merger and from the accretion disk left behind after the merger. Theoretical models predict that there will be considerable diversity in the electromagnetic counterparts to such events depending on the total mass of the binary neutron star system, the mass ratio, and the equation of state of dense nuclear matter, which sets the maximum mass of a neutron star beyond which it collapses to a black hole. Characterizing this diversity using observations of light and gravitational waves will be critical for unraveling the physics of accretion and jet production in binary mergers and their role in the origin of the heavy elements. Given the larger distance to gravitational wave sources as the facilities become more sensitive, this will require new observational capabilities for electromagnetic follow up. Large ground-based optical-infrared telescopes for spectroscopy to characterize heavy element production and sensitive ground based cm wavelength interferometers and space-based high energy telescopes to characterize relativistic jets will be particularly important. The science produced by joint gravitational wave and electromagnetic observations is likely to extend to at least a subset of LISA sources, namely binary black hole mergers in gas-rich galaxies. This will enable detailed studies of host galaxies, reveal the role of gas in facilitating massive black hole

mergers in galactic nuclei, and provide a sample of black hole mergers at high redshift suitable for cosmology.

Part of the power and promise of gravitational wave measurements is their ability to simultaneously enable "new astronomy" and "new physics." Tests of General Relativity using gravitational wave measurements are in their infancy. In the coming decades these tests will become far more stringent, with increasingly more precise measurements either cementing the quantitative applicability of General Relativity in the strong field regime, or perhaps revealing new physics. Sensitive ground-based gravitational wave detectors at higher frequencies and louder signals will constrain the radii of neutron stars through their tidal deformation. This in turn will constrain the equation of state of nuclear matter better than with current measurements, and in a way that is not possible in laboratories on Earth. Gravitational wave constraints on the neutron star equation of state will complement ongoing electromagnetic efforts using radio and X-ray timing and X-ray spectroscopy, likely leading to high precision measurements of neutron star radii in the coming decade. Larger samples of binary black hole and neutron star mergers will also significantly increase their utility for cosmology. Gravitational wave detections determine the distances to sources. Simultaneous electromagnetic observations of host galaxies to determine redshift can thus provide a measurement of the Hubble constant that is completely independent of current techniques.

Gravitational wave astronomy started with a bang in 2015, opening a new window to the universe with the detection of merging black holes by LIGO/Virgo. Over the few years since then, gravitational wave observations have become an indispensable astronomical tool. The coming decade, with the potential of detections in other parts of the gravitational wave spectrum, signals from new sources, and large numbers of black hole and neutron star detections, will be the start of a new era of precision and multi-wavelength gravitational wave astronomy.

2.2.5 Astronomical Transient Events

Although the night sky looks placid, millennia of observations have shown that it in fact varies systematically on many timescales. There are secular changes in the positions and speeds of objects due to their motion—be it interstellar interlopers in the solar system (see Section 2.1) or stars orbiting the 4 million solar mass black hole in the center of the Milky Way Galaxy. On top of this, there is a dizzying variety of dynamic astrophysical events that emit large amounts of energy in anywhere from the blink of an eye to timescales longer than human lifetimes. Some of these are cataclysmic events that herald the formation of neutron stars and black holes in stellar core-collapse, the thermonuclear explosions of white dwarfs, or the mergers of stars or compact objects. Others are repeating phenomena, such as stellar flares or the explosions of the surface layers of white dwarfs. All of these phenomena involve astronomical sources that produce light at many wavelengths, as well as in some cases gravitational waves and high energy neutrinos and cosmic rays. Astronomical transients impact nearly every area of astronomy. Supernovae and neutron star mergers disperse heavy elements into the interstellar medium, seeding gas with the elements necessary for life. Thermonuclear explosions of white dwarfs are valuable "standard candles" used to trace the acceleration of the universe. Fast radio bursts, millisecond bursts of radio emission of uncertain origin, have the potential to become a powerful new probe of the distribution of baryons throughout cosmic ecosystems (Section 2.3). Dedicated archives of brightness measurements now spanning more than 100 years enable the study of transient phenomena that recur on decade timescales or longer, such as stellar occultations by circumstellar material, the slow but dramatic brightening of newly unveiled young stars and of old stars like Betelgeuse in their death throes (see Section 2.1).

Rapid advances in detector technology and computing power have led to a revolution in astronomical time-domain surveys, which in turn have led to the discovery of new classes of transients (e.g., fast radio bursts and stellar mergers; Figure 2.16 depicts various classes of optical transients). The Rubin Observatory will take this effort to the next stage. Just one night of observation is expected to

detect 10 million transient events with sub-arcsecond positional accuracy, sending triggers to telescopes around the globe and in space, to follow up with observations at other wavelengths and to correlate with observations using other messengers. This revolution is extending to a broader range of wavelengths outside the traditional optical and gamma ray bands. The Roman satellite will carry out a near-infrared supernova survey that will also likely discover new classes of infrared transients. Roman's microlensing survey will allow characterization for the first time of the mass function of the majority of neutron stars and black holes in the galaxy. The extended Roentgen Survey with an Imaging Telescope Array (eROSITA) is carrying out the first all-sky X-ray survey since the Roentgen Satellite (ROSAT) in the early 1990s. Additionally, there is a tremendous increase in the number of international radio facilities searching for radio transients (e.g., CHIME, the Australian Square Kilometre Array Pathfinder (ASKAP), MeerKat).

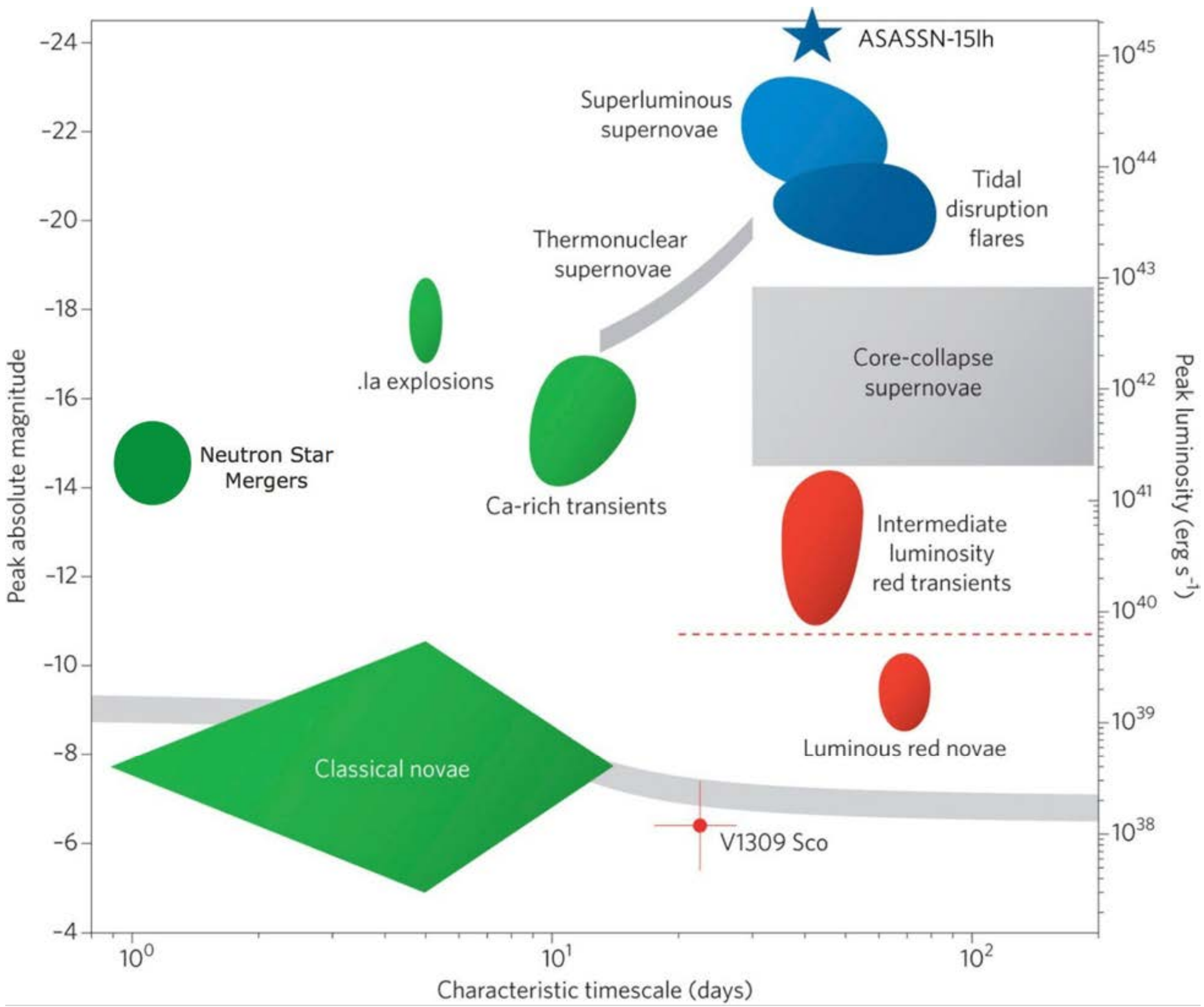


FIGURE 2.16 Types of optical transients, distributed in peak brightness and characteristic timescale, adapted from Cenko (2017). New wide-field optical surveys expected to come online this decade will surely populate this diagram with other transient exotica. SOURCE: Adapted by E. Quataert, with permission from Springer Nature: S.B. Cenko, 20187, Astrophysics: The true nature of transients, *Nature Astronomy* 1:0008, https://doi.org/10.1038/s41550-016-0008, copyright 2017.

The next decade has the potential to realize the full capabilities of transient and multi-messenger science, and very likely discover more surprises along the way. This work will ultimately provide a mapping between transients, the energy sources and central objects (e.g., black holes, neutron stars, stars) that power them, and their broader astronomical consequences (e.g., feedback and nucleosynthesis). Fully realizing this vision will, however, require a wide range of observational, theoretical, and software capabilities. The data required are obtained with simultaneous and coordinated operation of different instruments on Earth and in space. Ensuring easy user access to this wealth of data and the ability to cross-correlate multiple sources of data will maximize the science enabled by existing and planned facilities (see Sections 4.4 and 7.4.1). In addition, NASA's workhorse hard X-ray and gamma ray transient facilities (Swift and Fermi, respectively) are aging and their longevity is uncertain. Higher sensitivity all-sky monitoring of the high-energy sky, complemented by capabilities in the optical such as from Kepler and TESS, is a critical part of our vision for the next decade in transient and multi-messenger astronomy. Likewise, there are tremendous scientific opportunities for dedicated transient facilities at other wavelengths (e.g., the ultraviolet from space and radio on the ground) and dedicated spectroscopic follow up facilities, to complement the major U.S. investment in optical and near-infrared imaging surveys.

Priority Science Area: New Windows on the Dynamic Universe

The combination of new multi-messenger probes of astronomical phenomena with the maturation of time-domain observations opens up tremendous discovery spaces across nearly all areas of astrophysics. Within this discovery landscape, driven by improvements in gravitational wave and neutrino detection, and upcoming facilities such as the Rubin Observatory, one priority area stands out: the application of these new tools to the formation, evolution, and nature of compact stellar remnants such as white dwarfs, neutron stars, and black holes, as probed by the gravitational wave signatures of their mergers, together with rare explosive events that can be explored by the unique cadence and multi-color sensitivity of the Rubin Observatory. Sensitive observations of high-energy neutrinos and charged particles add new elements of discovery space, which will probe the universe's most extreme particle accelerators—New Windows on the Dynamic Universe.

The formation and evolution of compact stellar remnants, signaled by their accompanying multi-messenger transient phenomena are now serving to probe the range of neutron star and black hole masses in entirely new ways. These measurements provide information about the nature of matter under the extreme conditions that cannot be replicated in the laboratory, and about how the most extreme compact stellar remnants are formed and evolve. The mergers of neutron stars, uniquely observable at very early times through their gravitational wave signatures, can inform how elements such as gold and platinum are produced, which has been a mystery for many decades. Physical conditions near the surfaces and event horizons of neutron stars and black holes also represent extremes of matter, density, and gravitation, and serve as unique probes of fundamental physics. The coming revolution in temporal observations of the sky non-electromagnetic messengers, and through new observational cadences and spectroscopic follow-up across the electromagnetic spectrum is at the frontier of modern astrophysics.

As with the other priority science areas, progress will require coordinated advances in observations, experiments, and theory. The Vera Rubin Observatory and future large-area high cadence radio facilities will increase the numbers of variable and transient objects by orders of magnitude, with regular sampling over time. The Rubin Observatory will be unique for an optical time domain facility, as it will provide multiple optical colors at uniform cadence with unprecedented sensitivity. Advanced algorithms utilizing artificial intelligence algorithms to sift through Rubin Observatory's massive amounts of data will find the interesting outliers which will be heart of future progress and discovery. Current facilities such as the Chandra, SWIFT, and Fermi space observatories and ground-based radio and OIR observatories will play vital roles in such follow-up work. The full exploitation of this potential for

multi-wavelength time-domain observations will require maintaining and expanding these observatories. Such facilities need not be large or expensive but need to be optimized for the task, and this survey recommends the establishment of such facilities both for space (including needed replacement of capabilities currently handled by aging facilities) and on the ground. Since this science tends to be global by its very nature, international cooperation both in complementary facilities and data sharing would greatly enhance the scientific outputs from these investments.

The major new facilities recommended by this survey will all play important roles in extending the power of these capabilities in the future. Many of the visible counterparts of these sources (often originating at cosmological distances) are extremely faint, beyond the limits of present-day telescopes, but should be within reach of the next generation of ELTs. The relativistic outflows produced by these events often can be readily detected in the radio, and will be prime targets for next-generation facilities such as the ngVLA (recommended for design studies by this survey). Design studies for a next generation ground-based gravitational wave observatory will set the pathway towards a revolutionary new facility in future decades. X-ray observations—critical to fully understanding the physics of these phenomena—motivate the design and construction of new facilities ranging from the scale of Explorers to a future large mission.

Foundational research is also essential for maximizing progress in this field. Chief among these will be support for theoretical modeling and simulations of these highly relativistic and energetic phenomena, including numerical relativity to determine the nature of the gravitational signatures, plasma physics to understand the particle acceleration, and dynamical modeling to determine the populations that could lead to compact object mergers. Efficiently assimilating the large flows of data from the surveys, extracting the measurements, and interpreting the observations pose major challenges, but with benefits that will expand our astrophysical knowledge in entirely new ways.

2.3 COSMIC ECOSYSTEMS

Processes on a wide range of time and length scales together drive the formation, evolution, and interaction of the remarkable diversity of objects we observe, from exoplanets and stars to black holes and galaxies. A confluence of advances in theory, computational modeling, and observational capabilities expected in the next decade will transform our understanding by identifying the key mechanisms shaping this web of interconnected systems.

2.3.1 Overview

Arguably the single most important lesson in the last ~30 years of understanding the origin of structure in the universe is that it is not a one-way street, dictated solely by gravity from large scales to small. The formation of some of the smallest and densest objects in the universe, stars and massive black holes, dramatically alters how most other astronomical objects form, from planets and galaxies to stars and black holes themselves. Stars and black holes impact their surroundings through a broad set of energetic processes collectively known as feedback. These span an enormous range of time and length-scales, from gas as close as the planet-forming disk around a young star to as distant as in another galaxy millions of light years away.

Many aspects of star and galaxy formation can be viewed as a cosmic tug-of-war between feedback and gravitational collapse, as illustrated schematically in Figure 2.17. It is now known that the luminous bodies of galaxies, far from being disconnected from their surroundings, are part of a vast system that includes their surrounding circumgalactic medium out to intergalactic scales. Theory predicts that giant rivers of gas flow into galaxies, but most of the gas in galaxies is subsequently ejected back out into the circumgalactic medium by powerful galaxy-scale outflows. The flow of matter and energy throughout the entire system is likely responsible for the commonalities and differences among galaxies, but the details of how have been elusive. Likewise, the flow of matter and energy within a galaxy—again

due to the combined effects of gravity and feedback—determines the distribution of gas in the interstellar medium and where and how stellar and planetary systems form. The same flows depicted in Figure 2.17 also disburse the heavy elements produced by stellar processes, from the carbon in our bones to the rare-Earth metals in phones.

Understanding the interplay of gravitational and feedback-driven processes is challenging in part because it involves such a wide range of length and time-scales. In addition, much like understanding human health requires understanding how cells function, myriad small-scale physical processes regulate the flow of mass and energy illustrated in Figure 2.17, because they determine how gas cools, sheds its angular momentum, and mixes with other gas.

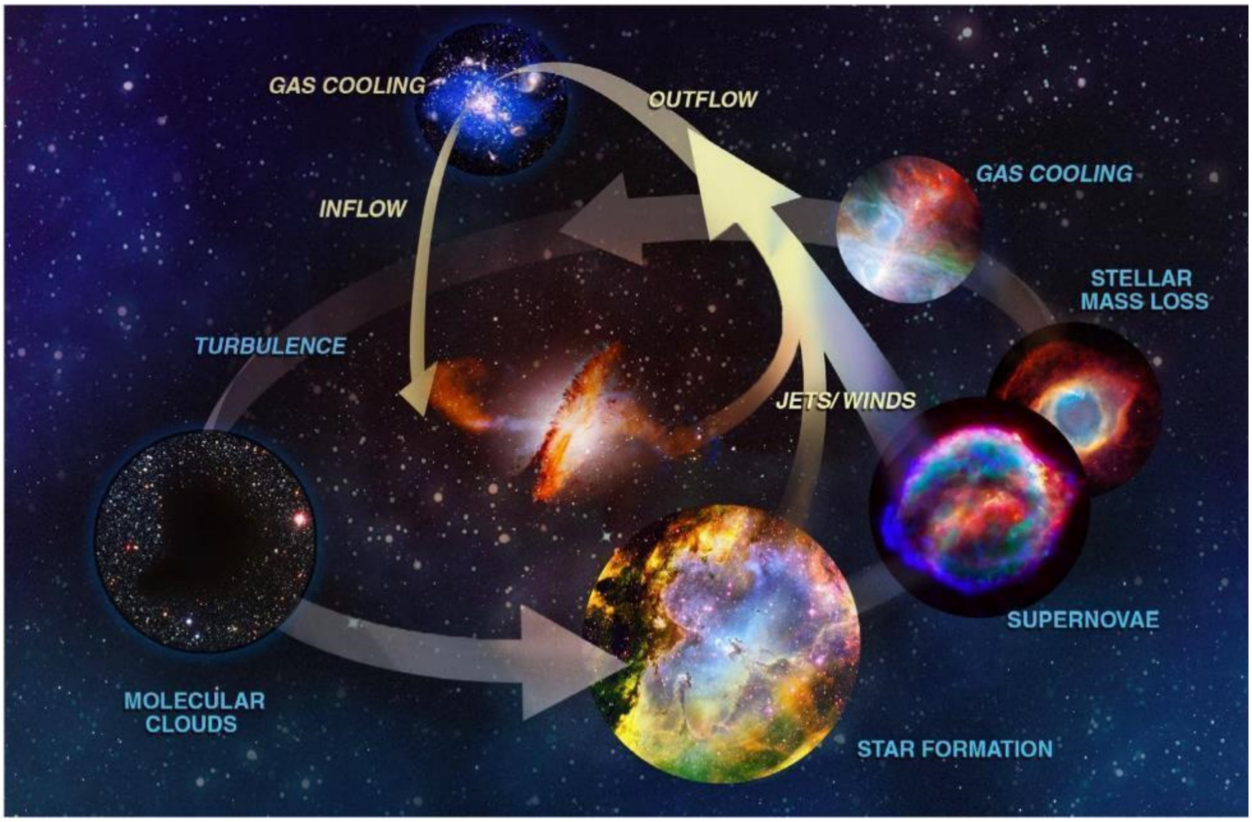


FIGURE 2.17 Illustration of the flow of gas into and out of the interstellar medium and galaxies through the combined effects of gravity and feedback. Heavy elements formed by fusion in massive stars are dispersed into the interstellar medium by stellar winds and supernovae. Much of this gas is in turn ejected from galaxies into the circumgalactic medium by galactic winds. Pristine gas accretes from the intergalactic medium to the circumgalactic medium, and subsequently accretes into galaxies, replenishing the fuel for star formation and subsequent generations of stars and supernovae. SOURCE: HABEX Report, The Habitable Exoplanet Observatory Study Team.

Small-scale physical processes at work in the cosmic ecosystem can thus have a surprisingly large impact on the large-scale behavior of astrophysical systems. An example is ionizing radiation, which is produced by massive stars and black hole accretion disks, and can regulate star formation and black hole accretion on sub-parsec scales. Some of this radiation can escape from the dense gaseous environments in which it was produced, and propagate out of galaxies into the intergalactic medium. In this way, the first stars and black holes were able to cause a global phase transition over scales of hundreds of Megaparsecs, in which most of the hydrogen in the universe was converted from a neutral to an ionized state, during what is referred to as the "Epoch of Reionization". Identifying the sources of cosmic reionization, and

better understanding how photons escape from their gas-rich and dust-enshrouded sources, will be an important science area over the coming decade. Numerical simulations of the propagation of radiation through galaxies (radiative transport) are highly computationally intensive, but critical for gaining a full picture of the role of this key process. Observational studies of gas kinematics, luminosity functions, and chemical compositions of galaxies spanning the Epoch of Reionization are needed, as well as studies of local galaxies that "leak" ionizing radiation, which may help shed light on the process of photon escape.

The symbiosis among cosmological phenomena on such different scales has been recognized for decades. Now, however, a confluence of advances in theory, computational modeling, and new observational capabilities will enable us to identify and understand the actual mechanisms at work in regulating this cosmic ecosystem. New observational probes of these interconnected flows at two widely separated spatial scales are highlighted in two of the discovery areas identified by the science panels: mapping the circumgalactic and intergalactic medium in emission and detecting and characterizing planets as they form.

2.3.2 Stellar and Black Hole Feedback

Stars have densities exceeding the mean density of the universe by more than 30 orders of magnitude; for black holes, the conditions are even more extreme. Despite their small sizes, stars and black holes are the most efficient sources of energy production yet discovered. This efficiency is a consequence of nuclear fusion in stellar interiors and the deep gravitational potential wells produced in stellar core-collapse and tapped by accretion onto black holes. Additional energy can be extracted at the end of a star's life. There is approximately one core-collapse supernova for every 100 M_{sun} of stars formed, and this single10^{51} erg supernova explosion can in principle accelerate 1000 M_{sun} of gas to speeds sufficient to unbind the gas from most galaxies. For accreting black holes the energy per unit mass is even higher. Thus, objects occupying a small fraction of the total mass and volume of the universe are critical to the evolution of much of the structure that is observed.

Many aspects of how stars live and die are currently uncertain enough that it limits the ability to model and interpret the effects of stellar feedback, be it in the form of radiation, stellar winds, or supernovae. Even in the local universe, for example, there are significant gaps in the understanding of stellar winds. This leads to large uncertainties in which massive stars become neutron stars and which become black holes. Stellar winds are also believed to play an important role in dispersing star-forming molecular clouds and driving turbulence in the interstellar medium, but the efficiency of this feedback again depends on the uncertain strength of winds from massive stars.

Even less is known about how the properties of stellar feedback vary with quantities like stellar metallicity, which strongly affects the efficiency of driving stellar winds, but varies from galaxy to galaxy and throughout cosmic time. Much of this uncertainty can be traced to the need for better theoretical models and observational diagnostics of stellar mass loss. More sensitive space-based ultraviolet (UV) spectroscopy is necessary to better characterize the spectra and winds of massive stars. Also, deep visible-wavelength and near-infrared spectroscopy of large numbers of stars in the Milky Way and other nearby galaxies from the ground—"industrial-scale" spectroscopy, one of the discovery areas identified by this Survey—would significantly sharpen constraints on stellar models.

Similarly large unknowns about feedback stem from uncertainties in binary stellar evolution. High mass binary stars evolve differently than single stars, which affects both their radiative output in life and the supernova explosions in which they die. Understanding binary stellar evolution is thus critical for understanding the global energetics of stellar feedback in galaxies. It is likewise important for understanding the spectra of galaxies, in particular the UV radiation that photoionizes the interstellar medium and likely reionized the universe during the initial epoch of galaxy and star formation. Separately, the dramatic advances in directly seeing the outcomes of binary stellar evolution with transient detection and gravitational wave facilities—for example, stellar mergers and compact object mergers—will continue to provide critical new insights into the life cycle of binary stars (see Section 2.2).

In addition to understanding the energy, mass, and momentum that stars supply to their surroundings, determining how these winds, radiation, and supernovae interact with the surrounding gas on different scales in the interstellar medium is equally important to untangling their interplay. Newly forming low-mass stars produce winds and jets that modify the structure of the clouds in which they form. They also produce radiation that heats and evaporates their surrounding protostellar disks, influencing the conditions for planet formation. Evaporation of planetary atmospheres by the same stellar radiation can also explain a bimodal distribution of planetary radii seen in transit observations (see Section 2.1). The higher energy radiation (UV and X ray) from low mass stars can also drastically alter the chemistry of planetary atmospheres, and thus the habitability of planets, but more precise determinations are needed of how the relevant radiation changes as a function of stellar mass or age.

In regions of high mass star formation, the radiation and stellar winds produced by massive stars can dominate the dispersal of molecular clouds (Figure 2.18), but observationally diagnosing which processes are the most important in different environments has proved challenging. There are tantalizing observational and theoretical clues suggesting that star-forming clouds with sufficiently high densities are difficult to disrupt by stellar feedback and may form super-star clusters and perhaps globular clusters at high redshift. Studies of reionization era galaxies (e.g., with JWST) and local analogues in the coming decade may finally resolve this long-standing puzzle. Regions of high mass star formation are often buried behind huge layers of dusty gas so improved long wavelength observations (far infrared, sub-mm, radio) are required to peer through the dust.

FIGURE 2.18 Multiwavelength image of the star-forming region 30 Doradus in the Large Magellanic Cloud illustrating the complex physical processes responsible for the disruption of star-forming giant molecular clouds and the production of hot, multiphase gas on the scale of 10’s of parsecs. Hot gas from stellar winds and supernovae (blue, Chandra), radiation from massive stars (green, Hubble), and re-radiated infrared emission from dust (red, Spitzer) all trace feedback into the interstellar medium on sub-kiloparsec scales. SOURCE: NASA, https://chandra.si.edu/photo/2012/30dor/, X-ray: NASA/CXC/PSU/L.Townsley et al.; Optical: NASA/STScI; Infrared: NASA/JPL/PSU/L.Townsley et al.

Supernovae generally explode too late after stars form to dominate the dynamics within most molecular clouds, but they are a critical source of feedback on galactic scales. The Chandra X-ray observatory led to major progress in the last two decades on understanding supernova feedback, but higher resolution and higher sensitivity X-ray imaging and spectroscopy would enable much more quantitative probes of supernova feedback and its role in powering galactic winds. Supernovae are important for a second, less direct, reason, in that they produce the gigaelectronvolt (GeV) cosmic rays

that dominate the energy of relativistic particles in many galaxies. The impact of cosmic rays is one of the largest uncertainties in understanding feedback in galaxy formation. The primary uncertainty is how cosmic rays are scattered by small-scale fluctuations in the magnetic field, which sets whether cosmic-rays can escape a region or whether their pressure builds up to the point where it can drive an outflow. On smaller scales, these cosmic rays can affect the thermal balance and chemistry of molecular clouds and their ability to form stars. It is remarkable that tiny solar-system scale fluctuations in the galactic magnetic field are a key ingredient in understanding how galaxies drive winds on scales of tens of kiloparsecs, or that the large scale magnetic field properties or distant supernovae can affect the formation of pre-stellar cores. This is an area where additional theoretical advances are particularly needed, including advances in plasma simulation techniques.

FIGURE 2.19 Multiwavelength image of the Perseus galaxy cluster illustrating the impact of black hole feedback in clusters and the coexistence of cool gas in the hot intracluster medium that provides fuel for ongoing star formation and black hole accretion. Radio emission from jets (pinkish lobes, Karl Jansky Very Large Array (JVLA)) fill cavities in the X-ray emission (violet, Chandra). Optical emission shows cooler photoionized gas (red filaments, HST). SOURCE: NASA and STScI, https://hubblesite.org/image/2376/gallery/135-multiwavelength, X-ray: NASA/CXC/IoA/A.Fabian et al.; Radio: NRAO/VLA/G. Taylor; Optical: NASA, ESA, the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration, and A. Fabian (Institute of Astronomy, University of Cambridge, UK).

Feedback from supermassive black holes during galaxy formation and evolution is one of the most dramatic examples of the small scales in the cosmic ecosystem impacting the large (Figure 2.19). Accreting black holes can influence their surroundings through UV and X-ray radiation, collimated relativistic jets, and wider-angle winds. However, the understanding of how active galactic nuclei (AGN) spectra, jets and winds vary with luminosity, black hole mass, black hole spin, and perhaps other properties is rudimentary. This currently is a major bottleneck in understanding the evolution of galaxies. Theoretical progress in these areas is likely to continue to be rapid, with advances in general relativistic simulations of black hole accretion predicting increasingly realistic jets and winds for comparison to observations. These predictions push the frontiers of radiation theory, plasma theory (to determine how and where the plasma is heated), and computational astrophysics. Observationally, a combination of high-resolution millimeter (mm) imaging and multiwavelength, multi-messenger observations can reveal the launching mechanism and particle content of relativistic jets (electron-proton vs. electron-positron) and thus the energetics of jet feedback. More detailed studies of molecular gas emission will reveal how the

densest star-forming components of the interstellar medium are impacted by AGN jets. And higher sensitivity and spectral resolution optical-UV and X-ray spectroscopy of broad emission and absorption line outflows from AGN are needed to better diagnose the physical properties of accretion disk winds and determine how much mass and energy they actually carry. In addition to measuring the jet and wind properties, better constraints on the masses and spins of the supermassive black holes themselves will play in important role in understanding how these objects formed and how they grow, as well as their role in the feedback processes described above. X-ray telescopes and next generation gravitational wave experiments can constrain the black hole spins, and their masses will be better constrained by measurements with next generation optical and radio telescopes.

Observations of galaxy clusters have revealed the critical role of black hole feedback by jets in the intracluster medium (Figure 2.19), though exactly how the energy from the black hole couples to the surrounding gas is still uncertain. This is a prime example of the need for multiwavelength observations: the combination of radio, X-ray, and optical data reveals the interplay between the centimeter-wave-emitting relativistic jets, mm-emitting molecular gas, X-ray emitting thermal intracluster plasma, and the optical-emitting photoionized gas. Higher spectral resolution X-ray spectroscopy of clusters would sharpen the understanding of AGN feedback in this critical environment that serves as a laboratory for understanding AGN feedback more broadly. Another observational frontier lies in extending studies of the hot intracluster medium to galaxies, probing the transition from galaxies that have largely ceased star formation to actively star forming galaxies. This is likely to transform the understanding of the role of feedback across a wide range of environments by directly observing the impact of feedback on the gaseous halos that contain the fuel for galaxy growth (see Box 2.3). Multiwavelength studies will again be key. UV and optical spectroscopy, and mm dust continuum observations probe the cooler multiphase gas. Combining UV absorption and UV emission studies of the CGM would be particularly valuable, as would deeper X-ray imaging and spectroscopy. The Sunyaev-Zel'dovich (SZ) effect is a direct probe of the thermal pressure of ionized gas in galactic halos. Of all the observational diagnostics at our disposal, the SZ effect most directly constrains the energy content of gas in galactic halos produced by the combined effects of gravitational collapse and stellar and black hole feedback. Higher sensitivity and higher resolution CMB observations motivated to a significant extent by cosmology (Section 2.2) will have a large impact on the understanding of galaxy formation as well.

2.3.3 Multi-Scale Cosmic Flows of Gas

Feedback and gravity are the key ingredients that determine how gas flows across cosmic scales. Directly observing these gas flows is challenging because of the diffuse nature of the gas in galaxy halos and the high spatial resolution required to peer into regions of ongoing star, planet, and massive black hole formation. New theoretical and observational capabilities are, however, allowing this critical aspect of the tug-of-war between gravity and feedback to be tackled.

Within galaxies, star-forming clouds form and disperse on timescales of millions of years. The structures are thus constantly being reshaped by cosmic flows of gas driven by the interplay between gravity and stellar feedback. Within those clouds, a major puzzle is how turbulence and magnetic fields determine the gas flows down to young stars and the planet-forming disks that surround them. Theoretically, this subtle problem requires understanding the degree to which the mostly neutral gas is coupled to the magnetic field—small-scale physics which dramatically impacts the large-scale problem of how disks around young stars form. Observationally, higher resolution radio and infrared imaging of protostars and their surrounding gaseous environments with ALMA and other instruments are required for progress in this area (Figure 2.20). Images of dust emission from ESA's Herschel Space Observatory revealed that gaseous filaments are responsible for fueling star formation on the scale of star clusters, but the role of filaments in determining cluster structure and stellar fragmentation is not yet clear. The accretion disks around young stars fed by these gas flows in star-forming cores set the conditions under which planets form.

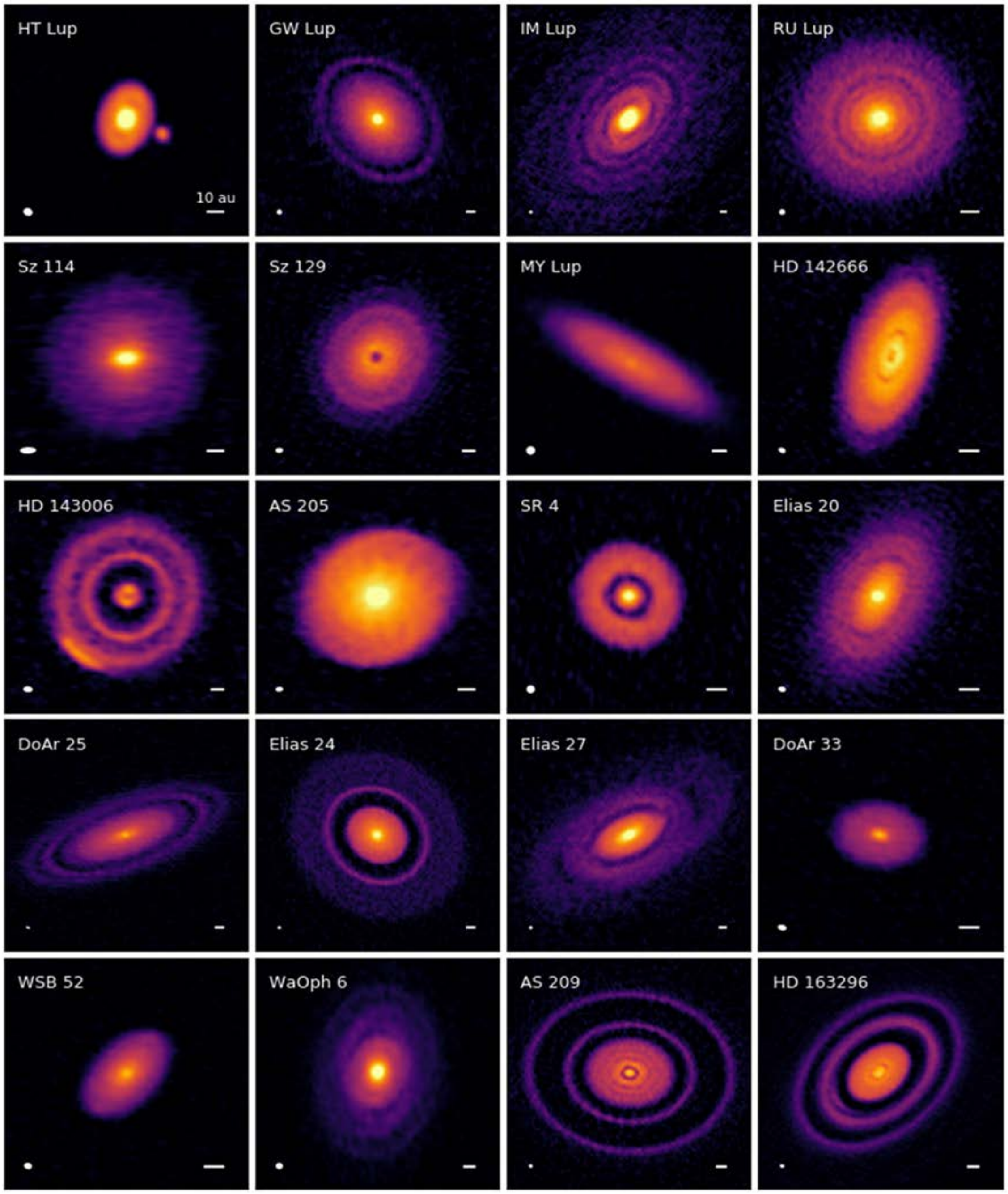


FIGURE 2.20 Montage of planet-forming disks around young stars as revealed by millimeter observations from the Atacama Large Millimeter/submillimeter Array (ALMA) observatory. ALMA's sensitivity and angular resolution enable the discovery of substructure in the disks as revealed by continuum dust emission. There is a near-ubiquitous geometry of gaps and rings which may point to the existence of forming planets, or possibly a signpost of magnetohydrodynamical processes occurring in the disks. Spectral-line observations indicate velocity structures in the gas and reveal flows similar to what has been predicted for the early stages of planet formation. In each image, the lower left icon indicates the beam size and the lower right icon is a 10 au scalebar. SOURCE: From of S.M. Andrews et al 2018, "The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, Sample, Calibration, and Overview," *The Astrophysical Journal Letters*, 869 L41. © AAS. Reproduced with permission. doi:10.3847/2041-8213/aaf741.

Studies with Spitzer, Herschel, and the Atacama Large Millimeter/submillimeter Array (ALMA) of local star-forming regions have identified protostars down to low masses and have brought the complex interactions with their surroundings to light. An area of particular focus is the relation between protostars and the dynamics, chemical evolution abundances, and physical conditions in protostellar disks. High resolution ALMA images of gas and dust have shown that many disks have small scale substructures such as gaps and rings (Figure 2.20), which are either created by already formed planets or by gas instabilities that could shape future planet formation. An unexpected result of these studies is that the central regions of the disks are often dust-obscured even at submillimeter wavelengths. Imaging of disks at high spatial resolution by ALMA or JWST can map the distribution within disks of a wide variety of molecules, with lines sampling an extensive range of density and temperature. These molecules, as interpreted by chemical models, are beginning to sketch out the early chemical evolution of planetary systems as shaped by the young star, but many key molecules, such as water, remain to be observed. A question for the coming decade is to understand the coupling between these small scale, solar system-forming regions, the larger cloud environment, and the diffuse ISM. The challenge is that these different scales harbor gas with a wide range of temperatures, phases, and chemical species, which require panchromatic observations from the millimeter and far-infrared through UV. Ground-based spectroscopy in the radio, optical, and near-infrared will be complemented by UV and far-IR spectroscopy from space to map the full cascade of star formation from the diffuse ISM down to individual protostars.

Our understanding of the flow of gas on the much larger distance scales of galaxies is also evolving rapidly. Despite the vast distances between observed galaxies, theoretical models predict that they are connected via the cosmic web of dark matter that forms the backbone for gas flows on cosmological scales (Figure 2.21). Theory predicts that gas flows into galaxies from the circumgalactic medium that fills the dark matter halos surrounding galaxies (Box 2.3). The circumgalactic medium is in turn fed by flows of gas from the more diffuse intergalactic medium and from gas ejected into the circumgalactic medium in galaxy-scale outflows. Characterizing the structure, metallicity, and dynamics of the circumgalactic and intergalactic medium with optical-UV and X-ray spectroscopy is a major frontier highlighted by the science panels.

During the formation of galaxies a small fraction ($\sim10^{-3}$) of the gas somehow loses essentially all of its angular momentum to end up in a massive black hole at the center of the galaxy. Exactly how this happens remains a mystery, though theoretical models are beginning to connect the growth of black holes to the properties of gas in the host galaxy on much larger scales (Figure 2.21). Given the energetic importance of black hole feedback, we need to understand how and when black holes grow, to assess their role in galaxy formation. High spatial and spectral resolution molecular gas observations are the key to peering into galactic nuclei, and can reveal how gas accretes and the properties of the black hole itself. One key problem highlighted by two of the science panels is whether massive black holes first form and grow by accretion from seed black holes formed in stellar core-collapse, or whether under rare circumstances it is possible for gas clouds to collapse directly to $\sim10^{4\text{-}5}\ M_{sun}$ black holes, perhaps via a supermassive star intermediary. Or does an entirely different set of processes seed galactic nuclei with massive black holes at high redshift? A combination of gravitational wave measurements of black hole mergers across cosmic time (see Section 2.2), deep near-infrared (JWST), far-infrared, and X-ray imaging and spectroscopy of high redshift accreting black holes would help piece together the origins story for massive black holes.

The collective result of feedback from star formation and black hole accretion powers galactic winds, galaxy-scale outflows of mass, energy, and heavy elements that have a major impact on the star formation histories and chemical evolution of galaxies and on the gaseous medium surrounding galaxies. The morphological and spectroscopic evidence for galactic winds is overwhelming (e.g., Figure 2.224), as is the heavy element enrichment they produce hundreds of kiloparsecs from star forming galaxies and even out into the diffuse intergalactic medium.

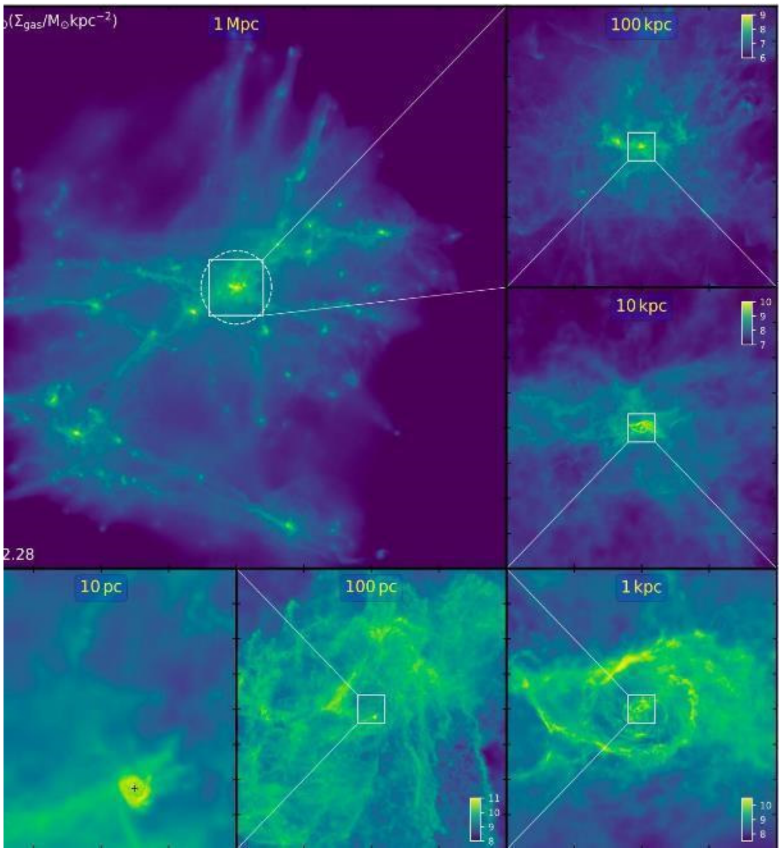


FIGURE 2.21 Flows of gas in a cosmological simulation from scales of the cosmic web on millions of parsecs to the central parsec of a galaxy, in order to study inflow of gas that fuels star formation and accretion onto the central massive black hole (the + in the lower left image). Color shows gas mass surface density. SOURCE: From D. Anglés-Alcázar et al 2021, “Cosmological Simulations of Quasar Fueling to Subparsec Scales Using Lagrangian Hyper-refinement,” *The Astrophysical Journal*, 917 53. © AAS. Reproduced with permission. doi:10.3847/1538-4357/ac09e8.

Massive stars and their supernovae are sources of mass, energy and momentum, which emerge in the form of fast-moving shock-heated gas, and relativistic cosmic ray particles as well as photons. Radiation, winds, and jets from accreting supermassive black holes are also likely important in powering galaxy-scale outflows in galaxies. It is, however, an open question whether AGN are only important for high mass galaxies or have a significant impact in low mass galaxies as well. There are major uncertainties in how the ingredients of stellar and AGN feedback combine to produce the outflows that we observe. Perhaps the biggest puzzle lies in understanding the subtle problem of the co-existence of gas at very different temperatures and densities in the outflows. Galactic winds are observed to be multiphase, with molecular, atomic, and warm and hot ionized gas apparently coexisting in the same outflow, but at different velocities. This is revealed by multiwavelength images and spectra ranging from the mm to the X-ray: the wide range of physical conditions requires a panchromatic observational approach. It is unclear how much of the colder gas is launched from the galaxy and how much forms in situ in the outflow. Recent far infrared dust polarization maps showing smooth vertical magnetic fields over large portions of galaxy disks argue that the cold gas in which this field is embedded is influenced by the wind; future infrared data will test these ideas and provide direct information on feedback processes in and around the cold ISM. Theoretically, it is likely that small-scale physical processes such as instabilities and thermal

conduction regulate the transfer of gas between different phases. Although they are small in scale, these processes are in fact critical because they determine how the wind material cools and thus how much mass and energy can actually be ejected from a galaxy. A better understanding of how mass moves between phases is also necessary to quantitatively interpret the wealth of multi-wavelength data on galactic winds, and connect those observations to the physical quantities of most interest such as the mass and energy winds carry. Some of the most pressing questions, and greatest opportunities, are on the largest scales, as outlined in Box 2.3.

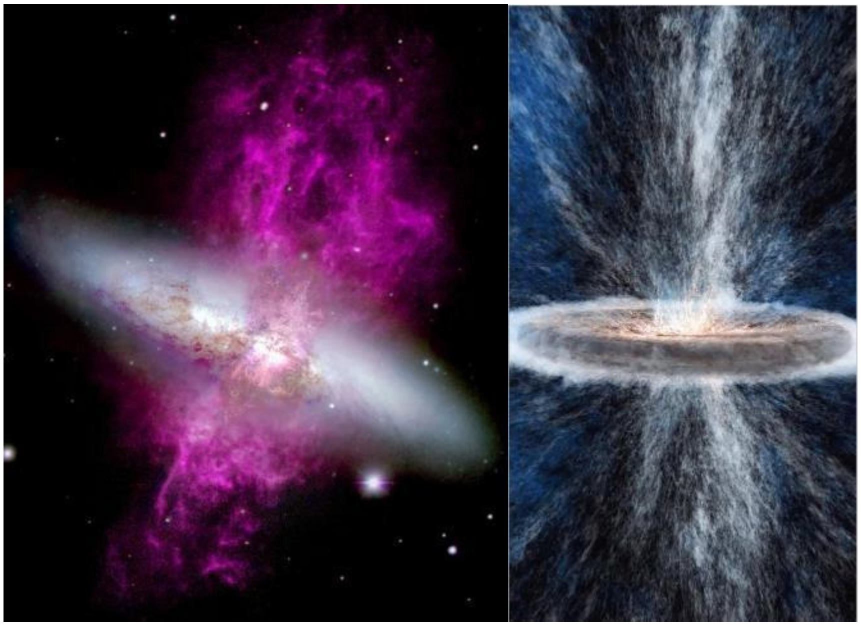


FIGURE 2.22 (*left*) Multiwavelength image of the starburst M82 illustrating the H-alpha emission from the galactic wind (pink) and the optical continuum from the galactic disk (BVI colors). (*right*) Numerical simulation of a bi-conical galaxy-scale outflow driven by multiple supernovae in a disk galaxy. SOURCE: Left: International Gemini Observatory/NOIRLab/NSF/AURA/M. Westmoquette (UCL)/J. Gallagher (Wisconsin-Madison)/ L. Smith (STScI/UCL). *Right:* From E.E. Schneider et al 2020, "The Physical Nature of Starburst-driven Galactic Outflows," The Astrophysical Journal, 895 43. © AAS. Reproduced with permission. doi:10.3847/1538-4357/ab8ae8.

BOX 2.3 Connecting Galaxies to the Cosmic Web

Observations of star-forming galaxies show that entire gaseous galactic disks would typically be converted into stars over less than a billion years, much less than the age of the universe at most redshifts. A continuous supply of gas into galaxies is thus required to maintain the observed ongoing star formation. This is believed to come from the circumgalactic medium (CGM) that fills the dark matter halos surrounding galaxies. The circumgalactic medium is in turn fed by flows of gas accreted from the more diffuse intergalactic medium and by material ejected from galaxies by galactic winds (Figure 2.3.1): the CGM is thus where all the elements of the galactic ecosystem connect. Based on constraints from current observations, about 10% of the baryons in the universe reside in stars and in the cold, dense interstellar medium, about 10% resides in the CGM within virialized galactic halos, and about 80% is in the diffuse medium. Within a virialized halo, the CGM may comprise 80% or more of the baryons.

Box 2.3 continued

All processes that drive galaxy evolution, from the evolution of individual stars and accretion onto black holes to the galactic-scale winds that are powered by them, take place against the backdrop of the large scale structure of the universe which feeds and sustains galaxies. This structure is gravitationally-dominated by dark matter, and simulations have long predicted that the dark matter is arranged into walls and filaments that are organized into the foam-like cosmic web. Gravitationally channeled by the dark matter filaments, gas is predicted to follow broadly the same large scale structure. Galaxies and the stars within them assemble at the nodes of the dark matter filaments, where gas accumulates, cools, and ultimately reaches high enough densities to form stars (Figure 2.21).

The interface between galaxies and the cosmic web is the circumgalactic medium (CGM), the complex and crucial immediate environment of galaxies. Giant rivers of pristine gas move inwards, from multiple directions, delivering the fuel for future generations of stars. These inflows are impeded as they encounter equally impressive galaxy-scale outflows, driven by star formation and black hole accretion deep inside the galaxies. This encounter creates shocks and turbulence, and it is now thought that the fate of galaxies—whether they continuously form stars at a low rate, like the Milky Way, undergo a period of intense star formation like M82 (Figure 2.22), or become quiescent—rests on the outcome of this interaction of competing gas flows. These interactions happen throughout cosmic history: the most intense winds likely occurred at "cosmic noon" ($z \sim 2$), with the CGM of today's galaxies providing a record of the past history of feedback. Furthermore, the flows are predicted to be strongly dependent on galaxy mass: theoretical models predict a transition from gas that is primarily hot and pressure-supported in massive galaxies to gas that is primarily cold in lower mass galaxies.

So far these spectacular interactions, which are among the largest causally-connected events in the universe, have almost exclusively been studied in computer simulations. Direct observations are needed to answer some of the most fundamental questions about galaxy formation, such as whether gas actually accretes from the intergalactic medium, what the rate of accretion is, whether gas in outflows is "recycled" back into the galaxies, and what the spatial distribution and physical conditions are of gas on the largest scales. Such observations also constitute one of the most direct tests of the long-standing idea that galaxies are connected with each other through a diffuse cosmic web of gas and dark matter.

The multiphase nature of the diffuse CGM/intracluster medium (ICM) and its complex dynamics, where many physical processes are superposed, necessitate a multi-wavelength and multi-scale approach. Improved theoretical studies of the key physical processes are needed, as are detailed computational models of global CGM dynamics faithful to this physics. On the observational side, major improvements are needed in imaging and spectroscopic studies of individual galaxy halos and clusters over the full spectral range from X ray to radio, and spatial scales that span from 100 kpc scale haloes down to 100 pc scale star-forming clouds. Very large optical telescopes with diameters of 20-30 m can probe the diffuse ionized gas directly with sensitive integral field units (IFUs) and other components through absorption line studies of background sources. A large telescope in space would be sensitive enough to allow ultraviolet absorption spectroscopy towards a dense network of faint background quasars behind a single foreground galaxy, revealing the composition, temperature, velocities, and density structure of the diffuse hot gas that is thought to contain most of the baryons. Large samples of dispersion measures to localized fast radio bursts (see Section 2.2.5) offer a powerful new probe of the ionized gas in galaxy halos. Together, these complementary observations would be transformational for probing and understanding the CGM. These are all major frontiers highlighted by the science panels.

Box 2.3 continued

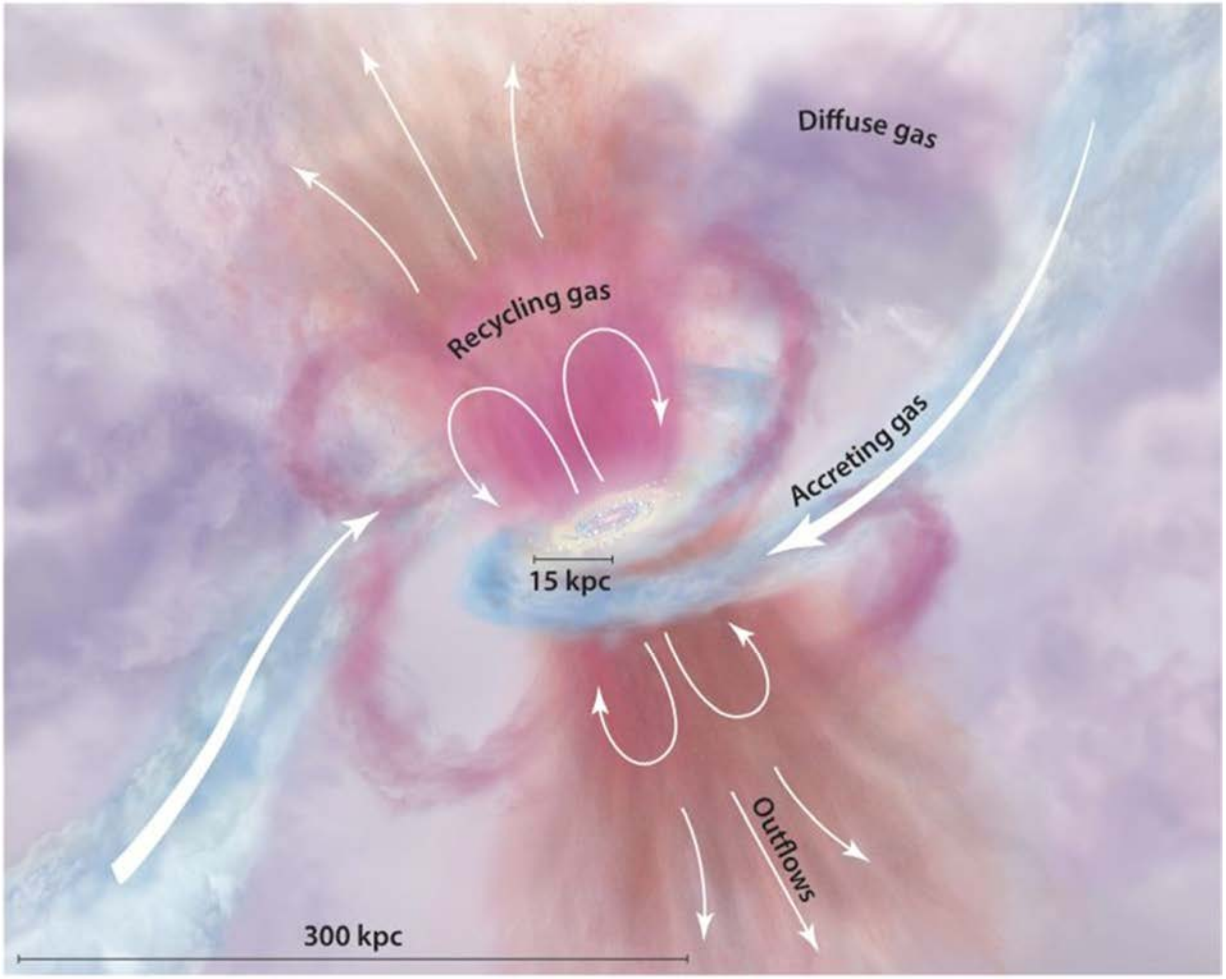


FIGURE 2.3.1 Illustration of the circumgalactic medium, where intense outflows from galaxies—driven by supernovae and black holes—interact with inflowing pristine gas from intergalactic space. SOURCE: Republished with permission of Annual Reviews, Inc., from Tumlinson, Peeples, and Werk, 2017, The circumgalactic medium, *Annual Review of Astronomy and Astrophysics* 55: 389-432; permission conveyed through Copyright Clearance Center, Inc.

Priority Science Area: Unveiling the Hidden Drivers of Galaxy Growth

By its very nature the *Cosmic Ecosystems* theme is very rich, exploring interconnected processes ranging over a billionfold range of linear scales, from individual supernovae to the virial radii of galaxies and beyond, Among this abundance of science objectives, the survey has chosen as its priority science area Unveiling the Drivers of Galaxy Growth. The pathway towards achieving this goal is bracketed by two sets of transformational observational opportunities: first the revolutionary measurements of the comic history of galaxy growth to come very soon from JWST, and culminating with the revolutionary capabilities for ground-based and space-based imaging and spectroscopy from the 24-40 m ELTs and a large IR/optical/UV space observatory, respectively.

An essential milestone on the pathway to revealing these drivers is a complete understanding of the formation and buildup of galaxies and their structures, stellar populations, metals, and central black

holes over cosmic time. JWST is poised to revolutionize this subject, by imaging and identifying galaxies out to some of the first generations during the epoch of reionization, and by obtaining deep redshifted ultraviolet – visible spectra for thousands of a galaxies spanning the period from reionization to the present. When combined with observations from ALMA, NOEMA, and the JVLA it should be possible to trace the transformation of baryons from interstellar molecules to stars and metals within a self-consistent framework. This information will be complemented by wide-field galaxy surveys from the Vera Rubin Observatory, the SPHEREx Explorer mission, the ESA Euclid mission, and the Roman Space Telescope, which will provide large statistical inventories of galaxies and rich target lists for future massively multiplexed spectroscopic surveys on the ground. Observations of nearby galaxies and our galaxy with JWST and WFIRST will also provide new insights into the physical processes driving star formation and the feedback processes powering the ecosystem.

Beyond that, the key missing link in unveiling the physics driving galaxy growth is to measure the properties of the diffuse gas within, surrounding, and between galaxies. These are the sites where baryons accrete on to galaxies, star formation is triggered, central black holes accrete and grow, and the feedback processes regulating galaxy growth are manifested. The key observational probes of all of these processes are emission and absorption-line spectroscopy of the diffuse gas, which contain a wealth of diagnostic information on the physical conditions, compositions, and dynamics of the gas. Current telescopes, whether located in space (HST) or on the ground (6-10 m apertures) are only capable of probing the densest regions in emission and occasional single sightlines through rare galaxies which happen to be superimposed in front of a bright quasar. These sporadic observations have been sufficient to demonstrate the power of spectroscopy for probing the baryon cycle but far too little to build up a robust physical picture of the processes at work. A large-aperture UV/optical space telescope, however, also envisaged for addressing Pathways to Habitable Worlds, would transform this subject. The combination of 6 m-class aperture and a high-efficiency spectrograph with modern detectors would provide thousands of potential sightlines to nearby galaxies, enabling “tomographic” studies of their circumgalactic and interstellar media, as well as rich new observations of the intergalactic gas clouds along the same lines of sight. With a versatility comparable to that of HST but an aperture comparable to JWST such an observatory would carry out groundbreaking observations of galaxy growth and address a majority of the science questions identified in the survey overall.

Major inroads to addressing this problem can also be met by the next generation of ELTs. This would include similar absorption-line spectroscopy (but targeted at the redshifted spectra of more distant galaxies observed at earlier cosmic epochs), as well as studies of the evolution of active galactic nuclei and the growth of the corresponding central black holes over cosmic time. The giant-aperture telescope alone will be able to obtain spectra for the faintest galaxies imaged by JWST, including some of the first generations of objects formed during reionization. The same spectroscopic capabilities, when trained on ancient stars in the halos of the Milky Way and in its nearest companion galaxies, can provide a detailed record of the assembly and chemical enrichment of these galaxies in their formative stages. More generally, the vast majority of faint galaxies revealed by deep JWST images will be too faint for follow-up spectroscopy either with the largest current ground-based telescopes or with JWST itself; the ELTs will play major roles in carrying out deep spectroscopic observations of these targets, probably including some of the most important “Rosetta stones” of the first generations of galaxies, to confirm their redshifts and measure their physical properties. Imaging observations with the ELTs (with AO) and with a IR/optical/UV space telescope of the resolved stellar populations of nearby galaxies out to the Virgo cluster can extend these fossil records to the full demographic population of galaxies today.

As with our other two priority science areas, these large facilities, though transformational in their potential impact, cannot address these critical scientific questions by themselves. Over the coming decade, major contributions to these investigations will come from multi-wavelength observations with the JVLA, ALMA, VRO, Euclid, Roman, 3.5-10 m OIR telescopes, Chandra, and toward the end of this period by the ESA Athena mission. Athena’s emphasis on *The Hot and Energetic Universe* will provide especially unique and powerful constraints on the cosmic feedback cycle, especially from AGNs, and it underscores the critical importance of future X-ray and infrared missions if these scientific questions are

to be fully addressed. Finally, perhaps as much as in any branch of astrophysics, progress in this field critically rests on a vibrant program of theoretical modelling and numerical simulations of galaxies, over the full dynamic range of scales over which the ecosystems operate.

2.4 SYNTHESIS AND CONNECTIONS

The science themes envisioned for the next decade and beyond reflect an interconnected cosmos. The hierarchies of structure inherent in the universe are dependent on interactions across this wide range of scales. Planet formation and evolution is influenced by the local stellar environment; star formation is affected by galactic environments and strongly influences them through supernovae and other feedback processes; and galaxy formation and evolution depends on the circumgalactic and intergalactic environment and interactions therein as well as on the energy output produced by massive black hole growth in galactic nuclei. Measurements of the positions and properties of individual stars in nearby galaxies enables an extraction of their evolution and life history from the larger forces at play in the galaxy's structure, formation, and evolution. Disentangling this web requires understanding the complex local interdependencies. The interconnectedness even extends to the power of applying multiple wavelength observations, multiple messengers, and novel observational techniques being applied to vastly different science areas.

The need for observations across the electromagnetic spectrum is common for all of these science themes, as the complexity of these questions requires a multiplexed effort. The interconnected nature of cosmic ecosystems necessitates an observational and theoretical program in which traditional boundaries between disciplines (e.g., radio, X-ray, star formation theory, galaxy formation theory) give way to a more comprehensive approach. Improved multi-wavelength observational capabilities are particularly important for probing the full range of physical conditions, from cold, dusty gas and synchrotron emitting cosmic-rays observed in the radio and infrared, to optical and UV observations of stellar and black hole radiation, to the hot interstellar and circumgalactic medium observed in the UV and X-ray and by the Sunyaev-Zeldovich effect on the CMB. The study of radio jets in galaxies informs the mechanical and radiative inputs to galaxy structure and details the large scale consequence of supermassive black holes at a galaxy's center, itself amenable to study at high energies, ultraviolet, optical and infrared wavelengths. Research into aspects of star formation requires a panchromatic approach: deeply embedded sources only appear in the infrared, while magnetically active young stars emit copious high energy radiation, and accretion processes appear at far infrared, optical, ultraviolet, and even high energy wavelengths. Compact objects can emit at the longest and shortest wavelengths of the electromagnetic spectrum, as well as potentially be a source of other particles like neutrinos or cosmic rays.

The universe is not static, and the time domain is important to many aspects of astrophysics. Constraints on the local value of the expansion of the universe rely on the identification and use of variable stars as one avenue for measurements, while a different route uses time delays in variability from multiple gravitationally lensed images of quasars to provide the constraint. Most of the detection methods for exoplanets use some means of detecting changes in stellar properties over time to infer the presence of a planet. Many objects in the universe change their intensity with time, in a way that illuminates the astrophysics of the object itself or makes it amenable to study some other phenomenon: witness the monitoring of quasar spectra over time to illuminate the innermost regions near the center of a galaxy around the central supermassive black hole, as well as studies of the lifetime and distribution of starspots through long-term precision light curves. The time domain also encompasses eruptions and explosions, from nova outbursts of episodic accretion to the merging of two neutron stars, which produce both electromagnetic and gravitational wave signals during the final coalescence, to the high energy flares that may prevent the development of life on otherwise Earth-like planets orbiting magnetically active stars.

In the same way that answering the science questions of the next decade requires a multi-wavelength and multi-messenger approach, these objectives also require the synergy of space, ground, and even underground facilities. In the gravitational wave arena, space-based detection of inspiralling

intermediate mass black hole mergers can signal the need for ground-based gravitational wave observation of the final merger, with searches for possible electromagnetic counterparts. Transients in electromagnetic emission will come by the millions per night from the Rubin Observatory once it is operational, with follow-up in other ground- and space-based telescopes needed for further characterization of the most interesting transient events. Neutrino observatories are opening an entirely new window for probing the most energetic processes in the universe. Exoplanet atmosphere characterization needs the combination of high precision spectro-photometric measurements obtainable from space with the high spectral resolving power available from large aperture ground facilities, to probe planets in the habitable zone around low-mass stars.

Theory, simulations, and laboratory measurements are just as crucial as new observations in making headway on these important lines of inquiry. A core challenge is understanding how to model systems in which processes on a wide range of time and length-scales interact to produce the universe that we observe. This includes both the need to model how the smallest objects (stars, supernovae, and black hole accretion disks) interact with their environment to understand the universe on large scales (e.g., galaxies and the circumgalactic medium), as well as the need to include small-scale physical processes (e.g., dust-gas instabilities in protostellar disks) to understand astronomical systems (planet formation, in this example). Theoretical mechanisms can explain the variety of transients observed and anticipated in the coming decade. General relativistic simulations of black hole accretion advance state-of-the-art predictions of the behavior of jets and winds from these compact objects, which can be compared with observations. From theory and simulations, a more complete knowledge of stars, as far as their rotation, binarity, and the impact of magnetic fields, improves the ability to model and interpret the expected ionizing output from massive stars. These parameters are also critical for understanding the evolution of massive stars and their feedback effects on galactic ecosystems. Likewise, simulations of convection, rotation, and magnetic field generation illuminate the dynamic nature of stars. Computational models and laboratory experiments, along with observations of gas and dust in planet-forming disks, are needed to understand planet formation in a more holistic manner. Knowledge of atomic and molecular properties gleaned from the laboratory can be essential for fully understanding the microscopic processes that have macroscopic consequences for stars, galaxies, and the cosmos.

TABLE 2.1 Science Panel Questions

Question	Theme(s)
Panel on Compact Objects and Energetic Phenomena	
What are the mass and spin distributions of neutron stars and stellar mass black holes?	New Messengers and New Physics
What powers the diversity of explosive phenomena across the electromagnetic spectrum?	New Messengers and New Physics
What do some compact objects eject material at nearly-light-speed jets, and what is that material made of?	New Messengers and New Physics
What seeds supermassive black holes and how to they grow?	New Messengers and New Physics, Cosmic Ecosystem
Panel on Cosmology	
What set the hot Big Bang in motion?	New Messengers and New Physics
What are the properties of dark matter and the dark sector?	New Messengers and New Physics
What physics drives the cosmic expansion and the large-scale evolution of the universe?	New Messengers and New Physics
How will measurements of gravitational waves reshape our cosmological view?	New Messengers and New Physics
Panel on Galaxies	
How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?	Cosmic Ecosystem
How do gas, metals, and dust flow into, through, and out of galaxies?	Cosmic Ecosystem
How do supermassive black holes form and how is their growth coupled to the evolution of their host galaxies?	Cosmic Ecosystem, New Messengers and New Physics
How do the histories of galaxies and their dark matter halos shape their observable properties?	Cosmic Ecosystem
Panel on Exoplanets, Astrobiology, and the Solar System	
What is the range of planetary system architectures, and is the configuration of the solar system common?	Worlds and Suns in Context
What are the properties of individual planets, and which processes lead to planetary diversity?	Worlds and Suns in Context
How do habitable environments arise and evolve within the context of their planetary systems?	Worlds and Suns in Context
How can signs of habitable life be identified and interpreted in the context of their planetary environments?	Worlds and Suns in Context
Panel on the Interstellar Medium and Star and Planet Formation	
How to star-forming structures arise from, and interact with, the diffuse ISM?	Cosmic Ecosystem
What regulates the structures and motions within molecular clouds?	Cosmic Ecosystem
How does gas flow from parsec scales down to protostars and disks?	Cosmic Ecosystem
Is planet formation fast or slow?	Worlds and Suns in Context, Cosmic Ecosystem
Panel on Stars, the Sun, and Stellar Populations	
What are the most extreme stars and stellar populations?	Worlds and Suns in Context, Cosmic Ecosystem
How does multiplicity affect the way a star lives and dies?	Worlds and Suns in Context
What would stars look like if we view them like we do the Sun?	Worlds and Suns in Context
How do the Sun and other stars create space weather?	Worlds and Suns in Context

TABLE 2.2 Science Panel Discovery Areas

Discovery Area	Theme(s)
Panel on Compact Objects and Energetic Phenomena:	New Messengers and New Physics
Transforming our View of the Universe by Combining New Information from Light, Particles, and Gravitational Waves	
Panel on Cosmology:	New Messengers and New Physics
The Dark Ages as a Cosmological Probe	
Panel on Galaxies:	Cosmic Ecosystem
Mapping the Circumgalactic Medium and Intergalactic Medium in Emission	
Panel on Exoplanets, Astrobiology, and the Solar System:	Worlds and Suns in Context
The Search for Life on Exoplanets	
Panel on the Interstellar Medium and Star and Planet Formation:	Worlds and Suns in Context
Detecting and Characterizing Forming Planets	
Panel on Stars, the Sun, and Stellar Populations:	Worlds and Suns in Context, Cosmic Ecosystem
"Industrial Scale" Spectroscopy	

3

The Profession and Its Societal Impacts: Gateways to Science, Pathways to Diversity, Equity, and Sustainability

"The pursuit of science, and scientific excellence, is inseparable from the humans who animate it."
—Panel on the State of the Profession and Societal Impacts

Every previous decadal survey of astronomy and astrophysics has stressed the importance of investing in people and has highlighted the value that astronomy and astrophysics brings to society, the nation, and the world. These investments and impacts have never been more important than today. The recent report *The Perils of Complacency: America at a Tipping Point in Science and Engineering* (2020) from the American Academy of Arts and Sciences urges dramatically increased investments in the preparation and diversity of future science, technology, engineering and mathematics (STEM) professionals to sustain U.S. scientific and technological leadership. This may be particularly important for the increasingly cyber future due to the need to understand and develop technology. These and other influential reports, such as the landmark National Academies reports *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (2007) and *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads* (2011), argue that increasing investment and diversity are needed more than ever to capitalize on multiple trends, including: the increasingly ambitious scope and scale of scientific research projects that rely on the creativity and capacity of the researchers and students who carry out the work; the increasingly global nature of scientific research, which increases competition for talent and innovation that require more attention to diversity and more expansive opportunity for participation; and the demands of policymakers and the public, whose investments are the primary funding sources for astronomy and astrophysics.

The Astro2020 decadal survey reflects the increased importance and attention on human investments and public impacts in multiple ways. First, the funding agency sponsors are increasingly visible and vocal on the urgent need to develop the nation's human capital, with a specific focus on what the National Science Board (2020) has termed "the missing millions" of individuals from traditionally underrepresented groups whose talent is needed for the success of the U.S. science and technology enterprise.[1] Second, the Astro2020 statement of task explicitly requires—as one of only five such explicit mandates—an assessment of, and recommendations pertaining to, the astronomy and astrophysics workforce and demographics. Finally, for the first time, the Astro2020 decadal process included a formal Panel on the State of the Profession and Societal Impacts (SoPSI; See Appendix N for the panel's full report).

This chapter necessarily distills the extensive documentation of the SoPSI report, and also considers some adjacent topics that were beyond the scope of the SoPSI statement of task; those additional topics, which are mainly discussed in Section 3.4, were taken up by working groups within the steering committee. Nonetheless, more than from any other single source, the contents of this chapter are informed and inspired by the SoPSI report, and by the diversity of voices and perspectives that it represents.

---

[1] National Science Board, 2020, *Vision 2030*, NSB-2020-15, Alexandria, VA, https://www.nsf.gov/nsb/publications/2020/nsb202015.pdf.

This chapter begins with a brief introduction of the key themes, including the precepts and principles that guide the ensuing findings, conclusions, and recommendations. Section 3.2 reviews the role that astronomy and astrophysics continues to play in creating novel technologies and providing crucial educational gateways to science, and discusses opportunities for increasing astronomy's impact on nurturing vital talent for the nation's global leadership in science and technology. Section 3.3 examines the factors that shape the current and future landscape of the astronomy and astrophysics profession, including the nature of the academic pipeline and the demographic makeup of the profession. Critical attention is paid to the ongoing need for efforts to make the profession more welcoming and inclusive, and more representative of the society to whom it is accountable. Then, Section 3.4 spotlights ways in which the future of astronomy and astrophysics necessarily depends on more sustainable practices in the utilization of and interactions with the world's natural resources, its cultures, and its human communities, including a major recommendation for the development of a new model for respectful, collaborative decision-making in partnership with Indigenous and other local communities. Finally, Section 3.5 summarizes the budgetary implications of our recommendations, and Section 3.6 concludes with closing thoughts. The central theme of people as a vital foundation will continue in the remaining chapters as well.

3.1 PRECEPTS AND PRINCIPLES FOR THE PROFESSION AND ITS SOCIETAL IMPACTS

The successful execution of the vision in this report will depend on the skill, creativity, and dedication of the community of scientists, engineers, educators, and aspirants who make up the astronomy and astrophysics profession. The ambitious facilities, instruments, and experiments envisaged by the Survey, and the transformative discoveries that they promise will not make themselves; the people who comprise the astronomy and astrophysics profession do these things. Because diversity of thought and perspectives fuels innovation, the astronomy and astrophysics enterprise can be at its most innovative only when it includes and embraces the diversity of its human talent, by ensuring equitable access to opportunities, eliminating barriers to participation, and valuing diverse forms of expertise in its leadership.

The societal benefits of investment in astronomy extend far beyond astronomy itself. As physical sciences, astronomy and astrophysics contribute to developing the nation's technically trained STEM workforce. Students with college-level training in astronomy and physics can access an extraordinarily broad range of technical careers—from education to national security to commercial R&D and beyond—that help fuel and sustain the nation's global leadership and well-being. Astronomical discoveries inspire people to pursue STEM careers generally, not only in astronomy. Impacting society even more broadly still, schoolchildren, teachers, parents, and the growing ranks of citizen-scientists benefit from opportunities for lifelong learning, analytical reasoning, and scientific literacy. Education has long been one of the great engines of social mobility. It is also a driver that transcends barriers, demolishes stereotypes, and unites those who offer or partake of it in a common purpose. In short, the astronomy and astrophysics enterprise adds substantial, real, and lasting value to the human knowledge infrastructure for the nation and the world.

Beyond these important tangible benefits, astronomy's quest to understand the universe and humanity's place within it resonates deeply with the public. Indeed, astronomy as a field is made possible because of taxpayers' and philanthropists' enthusiasm for the wonder and awe that astronomical discovery and achievement routinely delivers. The returns on national investment transcend the practical gains of STEM technology and workforce development by offering everyone the opportunity to experience the cosmos and to bear witness as astronomers unlock the answers to cosmic mysteries.

For these reasons, the nation's investment in astronomy and astrophysics as a science necessarily involves a substantial investment in people, both for the functioning of the field itself and for the many societal benefits that it produces. As with any investment, these investments in people require responsible

stewardship, and they demand transparency and accountability for outcomes, importantly through the collecting, evaluating, and acting upon reliable demographic and organizational data.

There is also the public's expectation that what is pursued with the nation's resources should be for the common good, which includes the principles of fairness and equal opportunity that are core to society's ideals. Not everyone can become a professional astronomer; but anyone with the ability and the drive to contribute to the nation through astronomical discovery should have a fair chance to do so. Astronomical activities also involve interactions among many peoples and countries of the world, and with Earth's climate and sky that all share; all would more greatly benefit from an engagement with astronomy that has sustainability as a core ideal. And everyone—regardless of identity or background—deserves the opportunity to bring their full true self to this enterprise free of fear, harassment, or discrimination.

The need to invest in people, and the potential outcomes for science and for the nation, have been called out by NSF and NASA as well. For example, the National Science Board's *Vision 2030* states that "the U.S. must offer individuals, from skilled technical workers to Ph.D.'s, on-ramps into STEM-capable jobs…In order to lead in 2030, the U.S. also must be aggressive about cultivating the fullness of the nation's domestic talent."[2] Similarly, NASA's *Science Plan 2020* states, "As research has shown, diversity is a key driver of innovation and more diverse organizations are more innovative . . . We will increase support by actively encouraging students and early career researchers. . . . We will also increase partnerships across institutions to provide additional opportunities for engagement and increasing diversity of thought. NASA believes in the importance of diverse and inclusive teams to tackle strategic problems and maximize scientific return."[3]

The precepts and principles articulated above—diversity, equity, benefit to the nation and the world, and sustainability and accountability—guide the recommendations that follow throughout this chapter.

3.2 ASTRONOMY'S ROLE IN SOCIETY: A GATEWAY TO STEM CAREERS, A BRIDGE BETWEEN SCIENCE AND THE PUBLIC

Astronomy, perhaps more than any science, has the power not only to educate but also to awe and inspire. Near-daily coverage of space science discoveries—images of the event horizon of a black hole, descriptions of exotic exoplanets—reveals the public's engagement with the field. For example, the August 21, 2017, solar eclipse was watched by an estimated 215 million Americans (two of every three people) either live or via videostream.[4] The Event Horizon Telescope image of the ring of light from plasma near the horizon of the black hole in the galaxy M87 posted on the NSF public website in 2019 was downloaded more times than any other image on a federal government server. The announcement of the detection of gravitational waves from a massive black hole binary by the LIGO-Virgo team in 2016 was the third highest-impact research story that year appearing in more than 900 news media outlets worldwide within one day of the announcement (Figure 3.1).[5,6]

---

[2] National Science Board, 2020, *Vision 2030*, NSB-2020-15, Alexandria, VA, https://www.nsf.gov/nsb/publications/2020/nsb202015.pdf.

[3] NASA, 2020, *Science 2020-2024: A Vision for Scientific Excellence,* NASA Headquarters, Washington, D.C., https://science.nasa.gov/science-red/s3fs-public/atoms/files/2020-2024_Science.pdf.

[4] J. D. Miller, https://isr.umich.edu/wp-content//2018/08/Final-Eclipse-Viewing-Report.pdf.

[5] See https://www.altmetric.com/top100/2016/.

[6] See https://www.aps.org/publications/apsnews/201608/backpage.cfm.

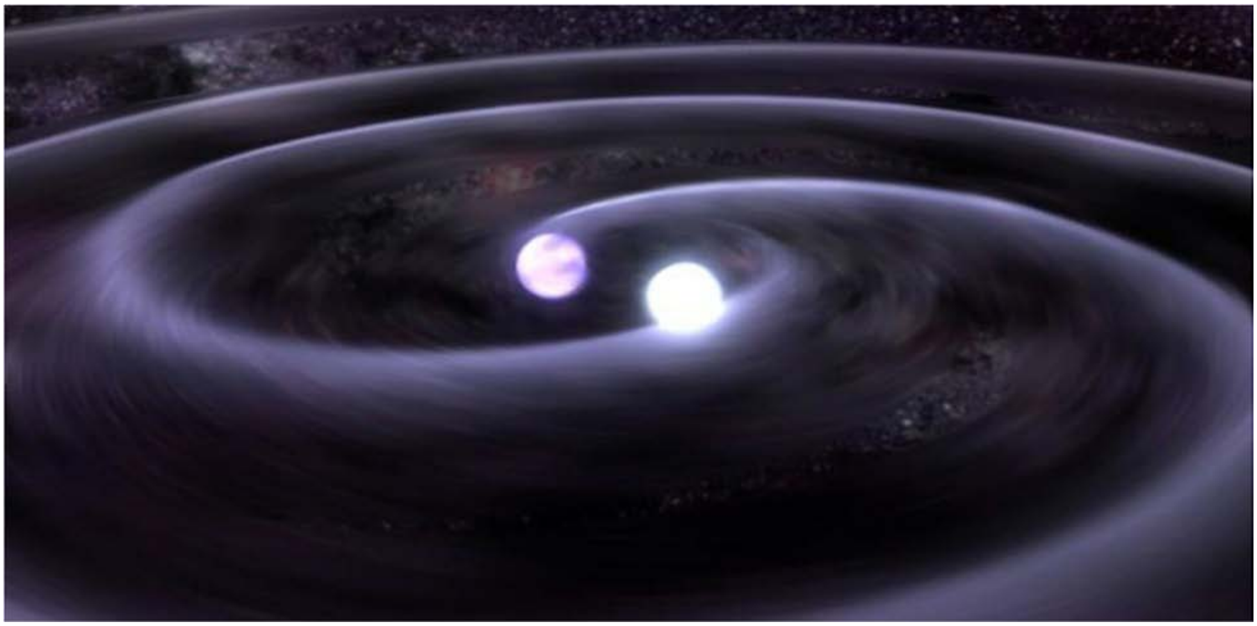


FIGURE 3.1 Black holes featured prominently in notable discoveries of the past decade that captivated the public's imagination. Artist's depiction of LIGO's first detection of gravitational waves from a black-hole merger event.[7] SOURCE: NASA/Dana Berry, Sky Works Digital.

Astronomical discoveries reach vast national and international audiences, and are often the first exposure that young people have to science and the scientific process. A small fraction of this audience will someday be inspired to take up a career in astronomy or space science, but for every one of those there are hundreds for whom the spark of an astronomical event or discovery will lead to a career in other areas of science, engineering, medicine, mathematics, computing, or technology. The term "gateway" is often used to describe this subject's ability to draw curious students to STEM. As counterpoint to a period when some have challenged the legitimacy of science and the integrity of scientists, the broad public appeal of astronomy can serve as a force for good far beyond the boundaries of its own discipline.

> **Conclusion:** Astronomy research continues to offer significant benefits to the nation beyond astronomical discoveries. These discoveries capture the public's attention, foster general science literacy and proficiency, promote public perception of the value, legitimacy, and integrity of science, and serve as an inspirational gateway to science, technology, engineering, and mathematics careers.

NASA, NSF, and the Astronomical Society of the Pacific have developed abundant K-12 and introductory college-level materials that are ready to bring astronomy into classrooms. These resources can impact the science literacy of millions of students across the country yearly. Indeed, a recent National Academies study examining the NASA Science Activation program for education and public outreach recommended that this high quality material should be made even more widely available and made readily accessible by K-12 teachers and college instructors.[8] The COVID-19 pandemic brought online education and digital learning resources into virtually every school and to every learner in the U.S., extending further still the opportunities for spreading astronomy educational materials across the country.

Astronomy is also a pioneer in developing "Citizen Science" projects such as the American

[7] See https://phys.org/news/2016-02-gravitational-discoveredtop-scientistsrespond.html.

[8] National Academies of Sciences, Engineering, and Medicine, 2020, *NASA's Science Activation Program: Achievements and Opportunities*, The National Academies Press, Washington, D.C:, https://doi.org/10.17226/25569.

Association of Variable Star Observers (AAVSO) and Galaxy Zoo,[9] which enable students and other members of the public to participate in scientific research, projects which have led to important new discoveries. Over the past decade more than 63,000 public volunteers from around the world have participated in programs run by the Zooniverse (Figure 3.2),[10] and the model has since spread to hundreds of other projects in the sciences, medicine, climate, arts, humanities, and social sciences.

> **Conclusion:** Astronomy is a leader in developing online citizen science projects, which enable students and other members of the public to participate in scientific research.

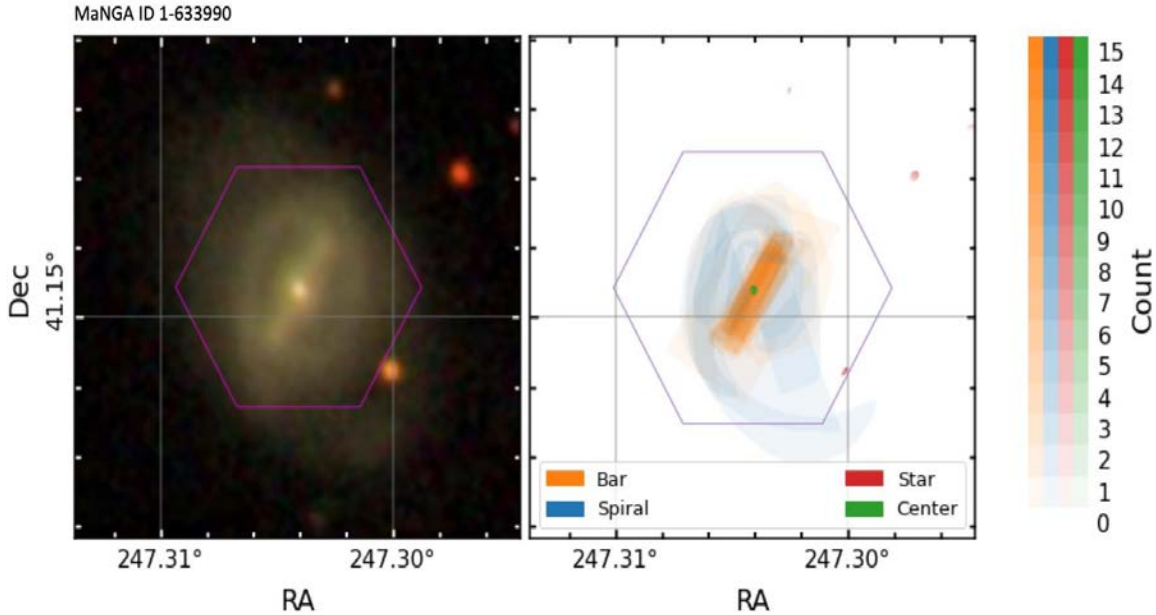


FIGURE 3.2 Astronomy has been at the forefront of citizen science, which has elevated traditional education and public outreach to true amateur-professional scientific collaboration. These images are from Galaxy Zoo 3D, a spin-off of the original successful Galaxy Zoo citizen science project. The image on the left shows a barred spiral galaxy with foreground stars; citizen science enables identification of substructures (as seen on the right), indicating where spiral arms, bars, and foreground stars are present in every galaxy observed by the SDSS IV MaNGA project. SOURCE: See https://blog.galaxyzoo.org/?_ga=2.141442354.1490305938.1619555147-768643056.1619555147. Adapted from Karen Masters (Haverford College), Coleman Krawczyk (University of Portsmouth). Galaxy image from SDSS. With thanks to Galaxy Zoo: 3D volunteers and the Zooniverse.org platform.

As a field that is driven by, and in turn drives, technological innovation, astronomy has always benefited the nation by invention and innovation of advanced technologies. In its essence, observational astronomy is remote sensing in the extreme. Its telescopes and instruments constantly push the limits of technology for precision and sensitivity, as they detect faint objects and extract delicate signals from a sea of noise. Its spectroscopy consists of detecting minute traces of chemical elements and molecules. Its reach extends from meter-length radio waves through the terahertz, infrared, visible, ultraviolet, X-rays, and gamma-ray parts of the electromagnetic spectrum, and has now extended even further to detecting

[9] See https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/.

[10] “Astro 2020 State of the Profession White Paper: EPO Vision, Needs, and Opportunities through Citizen Science” and “Astro 2020 Infrastructure Activity White Paper: Citizen Science as a Core Component of Research Infrastructure” by Laura Trouille (2020).

energetic cosmic rays, neutrinos, and gravitational radiation from sources billions or trillions of kilometers away. Many of these technologies, whether they be innovations in detectors, wireless communication, information technology, algorithms, or even in public engagement and communication, have propagated as "spin-offs" to other sectors of STEM and the commercial sector. What follows are just a few examples from the past two decades.

- The technical demands of NASA space missions have been especially productive incubators for spinoffs. (NASA has documented more than 1900 spinoffs since 1976.[11]) Some most closely tied to astronomical applications include complementary metal oxide semiconductor (CMOS) imaging sensors (used in most smartphone cameras today), infrared thermometers, and image enhancement and analysis systems. Technology sent to Mars for the first time on the Perseverance Rover is already detecting trace contaminants in pharmaceutical manufacturing, wastewater treatment, and other important operations on Earth.
- The demands of ground-based astronomy have provided a similarly rich harvest of technologies that have found widespread application in society, though the time for their adoption sometimes is measured in decades. These include early prototypes of WIFI, atomic clocks, cryogenic cooling systems (also developed by NASA for space missions), and the underlying technologies making possible precision location of 911 calls and (with significant additional investment from the military) GPS navigation.[12] The latter requires corrections for the influence of Earth's gravitational field on GPS signals, an unanticipated application of Einstein's theory of general relativity developed more than a century ago. GPS in its modern precision form would not function without these corrections. (Figure 3.3)
- Recent years have seen major improvements in the sensitivity of mm-wave and TeraHertz detectors. At the mm wavelengths, arrays of thousands of ultra-sensitive bolometric detectors have been developed to study the cosmic microwave background (CMB). In parallel, there has been steady improvement in radio-like receivers, but at a much shorter wavelength. These are exemplified by ALMA's Band 10 at 0.9 THz (roughly 0.3 mm wavelength) and, above 1.2 THz, by receivers based on hot electron bolometers. THz radiation can penetrate objects such as plastic and clothes, but not metals, and are not harmful to human tissues, and thus existing and in-progress sensitive detectors of THz signals have wide application in airport security and medicine. These developments parallel the history of X-ray technology, another spin-off from astronomy in the 1960s.
- Software and information technology are other areas where the footprints of astronomy have left clear marks. Grid computing is a prime example. The open source infrastructure "BOINC" developed in the Space Sciences Laboratory at the University of California at Berkeley for volunteer and grid computing was developed to search data obtained with radio-telescopes for signals from extraterrestrial life (SETI@home). It has since been used in many other areas in astrophysics (LIGO (+Virgo) application of BOINC is looking for evidence of continuous, monochromatic gravitational waves from non-axisymmetric, unknown single neutron stars in the Milky Way galaxy and LIGO noise diagnostics, for example) but also in many non-astronomical contexts including medical, environmental and humanitarian research sponsored by IBM Corporate Citizenship in the non-profit "world community grid", and even has been used for COVID-19 research. Extending upon this, training and collaboration with computing and data science researchers could be an additional area of broad benefit in the context of the cyber future.

[11] See https://spinoff.nasa.gov/.
[12] Radio Astronomy Contributing to American Competitiveness, NRAO/AUI report, 2006, https://www.nrao.edu/news/Technology_doc_final.pdf.

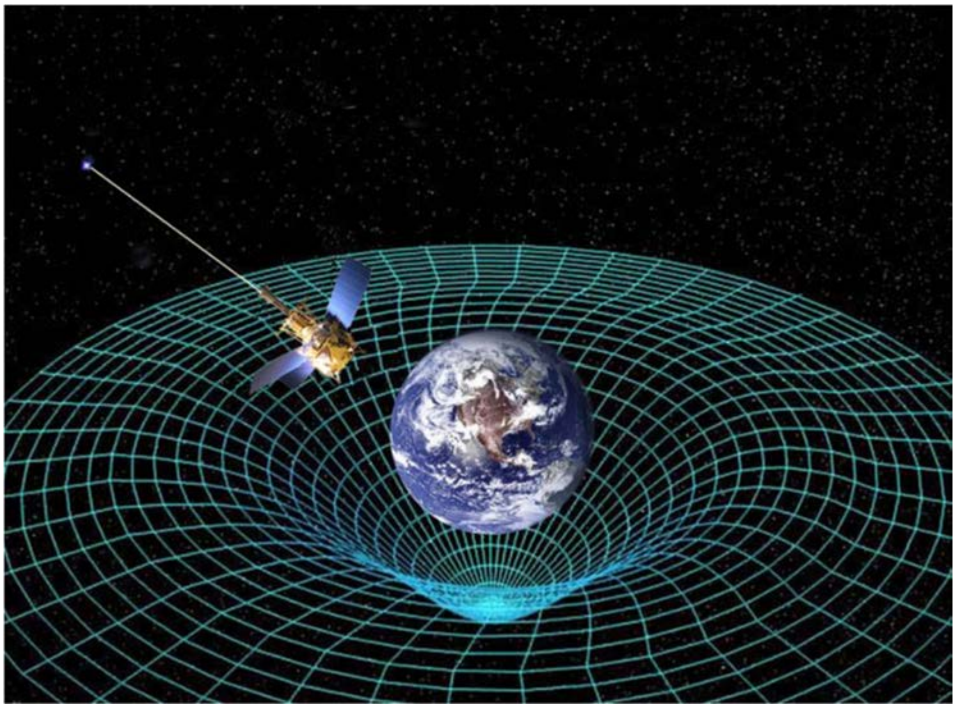


FIGURE 3.3 Astrophysical corrections accounting for the effects of general relativity on satellite timings are essential for high-precision GPS location and navigation services on Earth. SOURCE: NASA.

Conclusion: Astronomy continues to benefit the nation by invention and innovation of advanced technologies.

For all of these reasons, training in astronomy and astrophysics continues to pay dividends, whether individuals transition into long-term professional astronomy positions, STEM workforce roles in the private or public sector, or non-STEM related jobs. The 2017 NSF biennial survey of earned doctorates shows a less than 2 percent unemployment rate of individuals with an astronomy master's or Ph.D. degree.[13] Those joining the private sector with a bachelor's or Ph.D. earn a median starting income of $60,000 and $120,000, respectively.[14] A significant driver of these employment outcomes may be the increasing importance of computational skills and data science that are increasingly included in astronomy training and research. Indeed, these skills position individuals for opportunities in a variety of in-demand sectors, such as defense, health care, or commerce, as well as teaching in the education sector.

Finding: Education in astronomy research provides valuable training for a broad array of careers in STEM.

One key indicator of the value of astronomy research training beyond astronomy itself is the fraction of astronomy Ph.D. recipients who forgo postdoctoral positions -- traditionally the next step toward a permanent position in astronomy research -- in favor of non-academic STEM workforce jobs. As of the most recent survey in 2015-16, nearly half of new astronomy Ph.D. recipients were moving directly

[13] See https://www.nsf.gov/statistics/srvydoctoratework/. Similarly, the Bureau of Labor Statistics reports unemployment for life, physical, and social science occupations was about 2% in 2019-2020: https://www.bls.gov/cps/cpsaat25.pdf.

[14] Mulvey, P. and Pold, J., 2019, Astronomy Degree Recipients One Year After Degree, https://www.aip.org/statistics/reports/astronomy-degree-recipients-one-year-after-degree.

into private sector jobs.[15] This is a significant shift in career pathways for Ph.D.-trained astronomers in just over a decade; at the time of the previous decadal survey fewer than 30% of astronomy Ph.D. recipients were taking the straight-to-industry career pathway.[16]

These shifting patterns in career interests and outcomes may signal a healthy shift in attitudes and expectations about what constitutes a “successful” career for those with astronomy research training. Going back a decade further still to Astro2000 and prior decadal surveys,[17] the fact that a significant fraction of Ph.D. trained astronomers were not obtaining or choosing permanent positions in astronomy research was seen as a cause for consternation. The question was: did the “mismatch” between the number of astronomy Ph.D. recipients and the number of permanent astronomy research jobs imply a need for policies to limit the number of students admitted to Ph.D. programs? No such policies were implemented, and as noted above the number of students interested in astronomy has only continued to grow even though the number of permanent astronomy research positions has not grown apace. The net result is the significant increase noted above in the number of individuals successfully and lucratively taking their astronomy research training into a broad range of STEM careers. Astronomy is now contributing more broadly to the nation’s technically skilled workforce, and there is no evidence of any mismatch at all (see, e.g., the income and unemployment statistics noted above) between the number of trained astronomers and the number of desirable career routes for which those with technical training in astronomy find themselves in high demand.

Conclusion: There is no evidence of mismatch between the number of Ph.D.- or postdoc-trained astronomers and the broad array of desirable career pathways into the STEM workforce.

At the same time, this technical and career landscape is changing rapidly. To keep astronomers current and competitive for jobs in the public and private sectors, even more deliberate professional development will be needed, specifically with regards to the ever-growing importance of advanced computational skills.[18] The recent report from the Joint Task Force on Undergraduate Physics Programs recommends embedding computational training explicitly as part of the undergraduate curriculum, with at least one first-year computer course and one upper-level methods/statistics course, with an applied focus to physics and astronomy.[19] Early career data scientists, as well as early career instrumentalists, must also be nurtured and incentivized, as these skills represent evolving capabilities key to the future of astronomy and astrophysics.

Conclusion: One way to further enhance the competitiveness of physics and astronomy students for the broadest range of careers is to embed computational training in the undergraduate curriculum, with at least one course on programming, with a focus on applications to physics and astronomy.

Despite the strong career outcomes for students who have pursued education and research training in astronomy, the discipline underperforms relative to its potential for training an even larger number of college students for STEM careers. Of the ~70,000 new college freshmen each year in the U.S. who express an intent to major in physical sciences, only 10 percent overall—and only 4 percent of underrepresented minorities—ultimately complete a Physics/Astronomy degree (see Table 3.2, Section 3.3), choosing instead degrees in the life sciences or social sciences or in non-STEM fields altogether

---

[15] Heron, P and McNeil, L. 2016, A report by the Joint Task Force on Undergraduate Physics Programs, http://www.compadre.org/JTUPP/docs/J-Tupp_Report.pdf.

[16] See https://www.nap.edu/catalog/12951/new-worlds-new-horizons-in-astronomy-and-astrophysics.

[17] See https://www.nap.edu/read/9839/chapter/1.

[18] Huppenkothen, D. et al. 2018 PNAS September 4, 2018 115 (36) 8872-8877.

[19] Heron, P & McNeil, L. 2016, A report by the Joint Task Force on Undergraduate Physics Programs, http://www.compadre.org/JTUPP/docs/J-Tupp_Report.pdf.

(Figure 3.4).[20] In contrast, in the life sciences the retention rate is substantially higher, at ~50 percent.[21] When interpreting such statistics it is important to recognize that the undergraduate curriculum for astronomers, whether they pursue degrees in astronomy, physics, or both, is dominated by coursework in physics, As a result statistics for physics and astronomy undergraduate education are often aggregated. It also implies that improvements in the undergraduate component of the career pipeline for astronomers needs to be closely coordinated with like efforts in physics education.

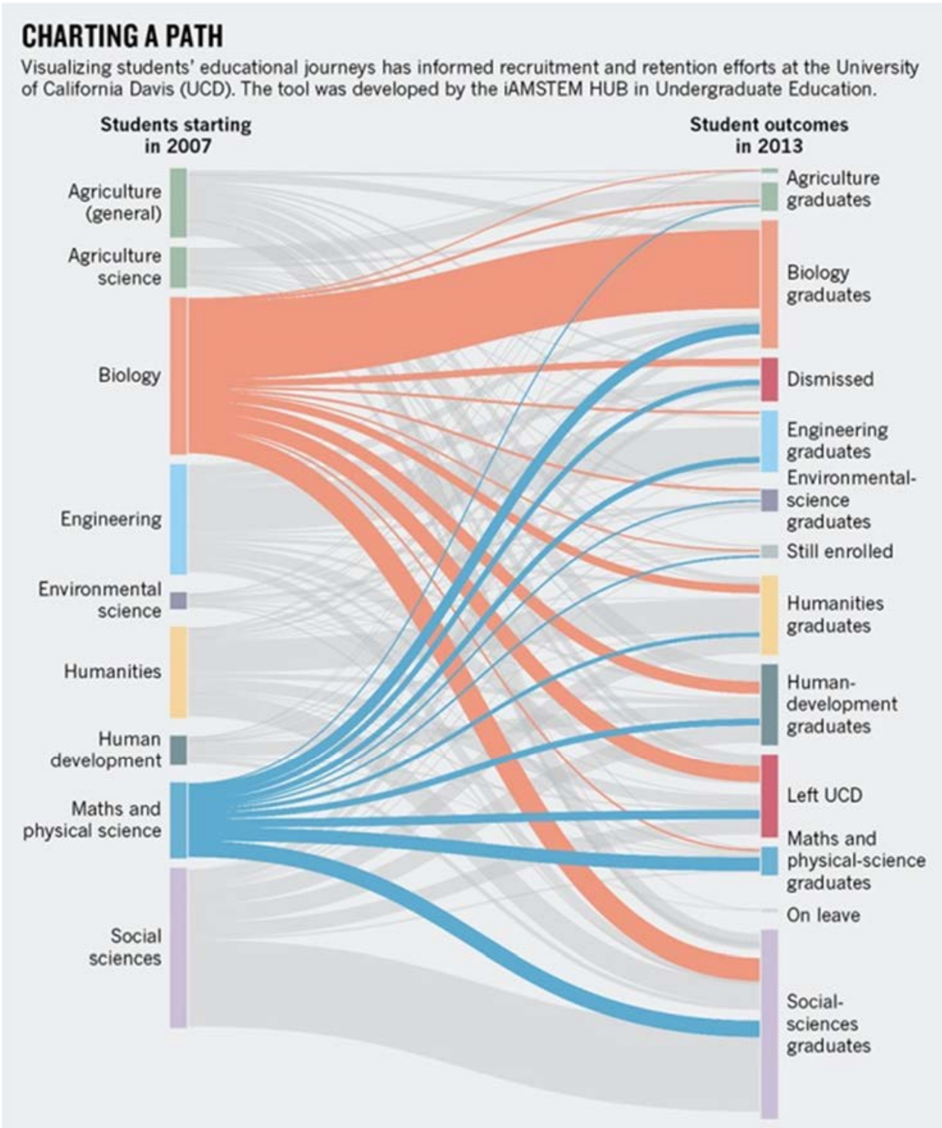


FIGURE 3.4 Results from a case study of undergraduate degree outcomes versus incoming student interests, by field, at the University of California Davis. In contrast to other STEM disciplines like biology or social sciences, physical sciences lose the vast majority of students arriving at college with an interest in those fields. SOURCE: Reprinted by permission from Springer Nature: S. Bradforth, E. Miller, W. Dichtel,, A.K. Leibovich, A.L. Feig, J.D. Martin, K.S. Bjorkman, Z.D. Schultz, and T.L. Smith, 2015, University learning: Improve undergraduate science education, Nature 523: 282-284, https://doi.org/10.1038/523282a, copyright 2015.

[20] https://www.nature.com/news/university-learning-improve-undergraduate-science-education-1.17954.
[21] Ibid.

Why do astronomy and physics capture such a relatively small market share of interested students? The answer, at least in part, could be that the (physics-dominated) curricula are aimed primarily at producing future academic leaders, often prizing the most basic and fundamental over the practical. As a result, students whose intellectual interests are in astronomy or physics, but whose practical career ambitions may lie outside of pure academic research, realize quickly that the curriculum and technical training opportunities are not intended for them. Indeed, quantitative and qualitative research of educational outcomes and student experiences consistently paint a very clear picture in which otherwise smart, capable students who could leverage their passion for astronomy and physics into meaningful STEM workforce careers not only choose to leave but feel "encouraged to leave."[22] This is in contrast to the messaging in many other disciplines, such as social sciences and biomedical sciences, which not only welcome and actively recruit interested students but intentionally structure the undergraduate curriculum and research training experiences at the undergraduate and graduate levels with the purpose of preparing the vast majority of students for successful careers outside of basic academic research.[23]

> **Finding:** The vast majority (>80 percent) of college students desiring technical careers and having an interest specifically in physics or astronomy, currently switch out of physics/astronomy and either obtain their technical training through another STEM field or else abandon STEM altogether, in contrast to the ~50 percent retention rate in the life sciences.

All of this suggests that astronomy and physics have a large opportunity to much more fully retain talented students and to much more fully contribute to the nation's technically trained STEM workforce, simply by shifting from a "weed out" mentality in the undergraduate curriculum, and from a "pure scientists only" mentality in research opportunities, toward approaches that much more intentionally attract and prepare—and value—students for the broad array of good career outcomes that astronomy and physics training provides anyway. The exclusive focus on academic careers, when options and positions are very limited, is overly constraining on trainees who might otherwise see industry as an interesting and lucrative career path through which they can continue to add value to the nation's technically skilled workforce. Indeed, the potential for advisors to guide students exclusively into academic careers and thereby discourage other good career outcomes, which may not serve the best interests of the students while simultaneously diminishing the overall STEM workforce pipeline.

> **Conclusion:** While astronomy and astrophysics has prepared students for a broad variety of technical careers in the public and private sectors, in practice advanced technical training in astronomy and astrophysics continues to largely select for those students most likely to seek academic research careers, representing a missed opportunity to welcome students interested in other applications and disciplines enabling astronomy and astrophysics to contribute more fully to the nation's broader STEM workforce pipeline.

3.3 FACTORS SHAPING ASTRONOMY'S CURRENT AND FUTURE PROFESSIONAL LANDSCAPE

As noted above, the astronomy and astrophysics profession is vital to the success of the Survey's vision specifically and of the astronomy and astrophysics enterprise more generally. A core principle and goal is to create an equitable field that allows full participation by all, and to achieve that goal requires identifying and addressing potential problems at every stage of training and practice. The SoPSI panel report provides extensive documentation and background references on the broad array of issues, challenges, opportunities, and potential solutions, the latter of which involve a combination of cultural change, removing structural barriers, and promoting accountability. This section briefly summarizes some

---

[22] https://www.springer.com/gp/book/9783030253035.
[23] https://www.nature.com/news/university-learning-improve-undergraduate-science-education-1.17954.

of the key issues and challenges, and distills the most pressing opportunities and solutions in order to provide guidance to the agency sponsors, policymakers, and the community. The focus here is primarily on areas that can be affected by agency funding, while acknowledging that this is only part of the larger work that needs to be done, and referring the reader to the full SoPSI report for details on additional areas of opportunity.

3.3.1 Where Astronomers Work

Almost everything about the way astronomers conduct their work—including the structure and size of research teams and the skill sets for which students are trained—has undergone massive shifts in the past two decades. The field is becoming dominated by large collaborations and survey-scale missions, an explosion of data, and a workforce that is more digitally connected and more geographically distributed than ever before. Indeed, occupationally speaking, astronomy research today bears little resemblance to the old stereotype of a lone scientist cloistered in a remote observatory. Rather, most astrophysicists' work has evolved to an "office job" over the decades, resembling in its rhythms, structures, and interactions the activities of most other modern-day white-collar professions. This includes an ever-growing recognition of the importance of—and expectation for—professional conduct (e.g., workplaces free of sexual harassment), professional development (e.g., intentional training for important technical, management, and leadership skills), professional work-life balance (e.g., accommodating the realities of childcare, eldercare, and other personal obligations), and other features that continue to make the astronomy profession, simply put, more professional.

According to a survey of American Astronomical Society (AAS) members (Table 3.1), more than half of full-time employed members of the profession with astronomy and astrophysics Ph.D.'s work at institutions of higher education; 33 percent work at government labs, research institutes, or observatories; and a few percent work in industry.[24] This pattern of employment and funding has held relatively stable over the past decade.

One consequence of this pattern of employment is that a large fraction of professional astronomers depend to varying degrees—in some cases to a large degree—on federal grant resources for their own support and/or for that of their research teams (Table 3.1). Another consequence is that, since the vast majority of astronomers' employers are divided between universities/colleges on the one hand versus large research centers/facilities on the other hand, the organizational approaches to workforce development may differ depending on organizational mission, structures, and mechanisms for accountability. For example, higher education institutions generally have teaching and training as core parts of their organizational mission, with accountability to parents, alumni, state legislatures in some cases, and university boards and leadership. And because they depend on federal funding for much of their research activities, the policies and priorities of these organizations can be influenced by the expectations and requirements of the funding agencies. In contrast, nearly all of the major facilities supported by NSF and NASA are operated through cooperative agreements, contracts, or other instruments with managing organizations (AURA, AUI, and others). It is not clear what accountability mechanisms the funding agencies have implemented with these organizations specifically with regards to training and employment outcomes.

These differences in employment contexts have implications as well for approaches to diversity and inclusion efforts. Many, though certainly not all institutions of higher education have implemented efforts toward greater diversity and inclusion as core elements of the organizational mission, and university-based investigators applying for federal grants are now routinely expected to address requirements for broader impact in their funding proposals, including with regards to broadening participation of underrepresented groups. Again, it is not clear what accountability mechanisms the funding agencies have implemented for the facility-managing organizations with regards to diversity and

[24] See https://aas.org/sites/default/files/2019-10/AAS-Members-Workforce-Survey-final.pdf.

inclusion expectations. However, the managing organizations have communicated a positive stance toward diversity and inclusion, with official policies, and with officials assigned to provide oversight, internal diversity and inclusion training, and promote community values. Most of NASA's research centers are managed by the agency directly, and thus NASA could in principle directly implement targeted procedures and accountability for outcomes. In addition, the increasing complexity of new observatories and observational methods can and has been attracting people from other engineering and science fields into important roles in astronomy; the excitement of astronomy can potentially draw in a wider and eventually diverse pool of engineers and other scientists.

TABLE 3.1 American Astronomical Society 2018 Survey of Employment and Salaries of AAS Members

Current Employer of US AAS Members with PhDs, 2018		
Employer or Sector	**%**	**N**
University or 4-year college	54	514
Govt. lab or research facility	14	135
Research institute	10	94
Observatory	9	86
Industry	4	39
Other govt.	2	20
2-year college	2	15
Self-employed	1	7
Planetarium or museum	1	7
Secondary school	-	4
Other	3	27
Total		*948*

Includes full-time employed respondents with PhDs excluding current postdocs.

Funding Sources for Salaries of US AAS Members 2018		
	% Receiving Funding	**Average % of Total Funding**
College/University	44	90
NASA	39	74
NSF	16	57
DOE	4	70
DOD	4	71
Foundation/Grant/Donors	3	74
Other	12	83
Total N		*1410*

Categories with <3% are not included

NOTE: These data represent only those individuals with active AAS membership; not reflected in these statistics are the large number of individuals who obtain academic degrees in astronomy and astrophysics but who "leave the profession" for jobs in the private or public sectors, and for whom the data suggest their training has enabled gainful employment in the STEM workforce (see Section 3.2). In the table at right, the rightmost column gives the percentage of a typical individual's salary that derives from a given source; for example, 44% of AAS members receive salary support through their college/university employer, and those individuals typically receive 90% of their salary support from that source. The "Total N" indicates the total number of people included in the survey; it is not the sum of the rightmost column as the formatting might suggest. SOURCE: https://aas.org/sites/default/files/2019-10/AAS-Members-Workforce-Survey-final.pdf.

3.3.2 Demographics of the Astronomy and Astrophysics Profession

The current demographics of the field, and trends in these demographics over the past decades, tell a mixed story. For example, with regards to gender Figure 3.5 indicates that the field still has a ways to go to achieve the higher levels of gender parity that are now the case in other physical science disciplines such as chemistry. At the same time, astronomy has now reached an important milestone in terms of gender representation, with the rate of PhD attainment among women now matching the rate with which women earn baccalaureate degrees (Figure 3.5). Indeed, as a discipline that is respected and influential in public opinion, astronomy's ability to model growth toward equitable participation and inclusive practices may influence other sciences and professions.

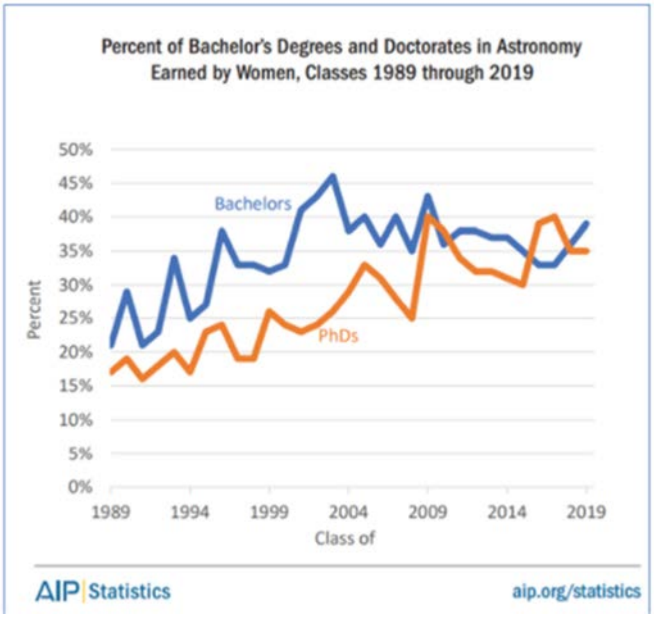


FIGURE 3.5 The percent of bachelor's and doctoral degrees earned by women in astronomy. The figure is intended to indicate the trend, keeping in mind that astronomy Ph.D.s also come from other undergraduate majors (e.g. physics). SOURCE: https://www.aip.org/statistics/reports/women-physics-and-astronomy-2019. Courtesy of Nicholson, S., and Mulvey, P.J., 2021, "Roster of Astronomy Departments with Enrollment and Degree Data, 2020: Results from the 2020 Survey of Enrollments and Degrees" © Statistical Research Center at the American Institute of Physics.

In addition, according to statistics from the American Institute of Physics (AIP),[25] the representation of women among the astronomy faculties of colleges and universities has shown clear improvement over the past decade, particularly among the recently hired assistant professors and recently tenured associate professors, for whom women now comprise about 30 percent, up from about 20 percent in 2003. There is a marked drop-off by roughly a factor of 2 in representation from the associate to the full professor ranks, though the absolute percentage of female full professors has increased to 15 percent (from roughly 10 percent) over the same period. At the senior ranks, the lower percentage of female faculty is in part shaped by lower fractions of women in Ph.D. programs in the past. In addition, AIP surveys show that women remain systematically disadvantaged by gender-associated differences in the distribution of family work and in career-advancing opportunities and resources[26] that may have become exacerbated by the COVID-19 pandemic.

> **Conclusion:** Ensuring the movement of women into the top leadership ranks (full professor and beyond) continues to be an important area needing attention.

Racial/ethnic diversity among astronomy faculty remains, in a word, abysmal. African Americans and Hispanics comprise 1 and 3 percent of the faculty, respectively.[27] This collective representation of 4 percent is about an order of magnitude below these groups' joint representation in the U.S. population. This underrepresentation was identified as a problem as far back as the 1980 Decadal survey.[28] As of

---

[25] Pold, J. & Ivie, R., Workforce Survey of 2018 U.S. AAS Members Summary of Results, https://aas.org/sites/default/files/2019-10/AAS-Members-Workforce-Survey-final.pdf.

[26] Porter, A.M., & Ivie, R. (2019) Women in Physics and Astronomy, https://www.aip.org/statistics/reports/women-physics-and-astronomy-2019.

[27] AIP Academic Workforce Survey, 2016, unpublished results.

[28] 1980 Decadal Survey (Field et al), v1 Appendix B (p172), and Vol. 2 starting on p. 334.

2012 there was not a single astronomy department that had representation of both African American and Hispanic faculty, and roughly two-thirds of astronomy departments had representation of neither. Funding agencies have traditionally invested in early-career faculty through dedicated programs such as the NSF CAREER awards and programs that support intentional transitions of postdoctoral researchers into faculty positions such as NSF Alliances for Graduate Education and the Professoriate (AGEP);[29,30] these can be valuable levers for incentivizing faculty hiring in general and, to the extent that such programs include diversity efforts in their selection criteria, can help to incentivize faculty diversity as well.[31]

> **Conclusion:** Racial/ethnic diversity among astronomy faculty remains abysmal. African Americans comprise a mere one percent of the faculty, over all ranks, among astronomy departments; Hispanics comprise three percent. This collective representation of four percent is roughly an order of magnitude below these groups' joint representation in the U.S. population.

> **Recommendation: Funding agencies should increase funding incentives for improving diversity among the college/university astronomy and astrophysics faculty—for example, by increasing the number of awards that invest in the development and retention of early-career faculty and other activities for members of underrepresented groups.**

3.3.3 The Academic Pipeline into the Profession

The past decade saw a substantial growth in the desire of Americans to participate in the excitement of astronomical discovery. The number of astronomy B.S. and Ph.D. degrees shows continued growth (Figure 3.6). There has been a steady increase in the number of women and Hispanic Americans earning astronomy degrees (Figures 3.6 and 3.7), though the number of African-Americans earning Ph.D.'s remains low and unchanged over three decades. Encouragingly, the number of African-Americans earning B.S. degrees has increased in recent years (Figure 3.7), making it all the more important to redouble efforts to recruit and support these students as they move into doctoral programs. Research suggests at least two key innovations in graduate STEM training, discussed below, that can help to address the persistent challenge of underrepresentation: (1) graduate training that is more explicitly motivated by pro-social concerns (i.e., work that is seen as positively impacting one's own communities),[32] and (2) more holistic approaches to evaluating individuals for entry to graduate programs.[33]

> **Finding:** The number of students pursuing undergraduate and graduate degrees in physics and astronomy continues to grow, and the field is becoming more representative of American demographics, with steady increases in the number of women and Hispanic Americans. Representation of African-American students, however, remains nearly steady and alarmingly low.

---

[29] https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503214.

[30] https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5474.

[31] Brown-Glaude, W. (Ed.). (2009). Doing Diversity in Higher Education: Faculty Leaders Share Challenges and Strategies. New Brunswick, New Jersey: Rutgers University Press.

[32] Jackson MC, Galvez G, Landa I, Buonora P, Thoman DB. Science That Matters: The Importance of a Cultural Connection in Underrepresented Students' Science Pursuit. Gibbs K, ed. CBE Life Sciences Education. 2016;15(3):ar42. doi:10.1187/cbe.16-01-0067.

[33] Innovation in Graduate Admissions through Holistic Review. Holistic Review in Graduate Admissions: A Report from the Council of Graduate Schools. CGS Webinar: Holistic Review in Graduate Admissions (February 2016).

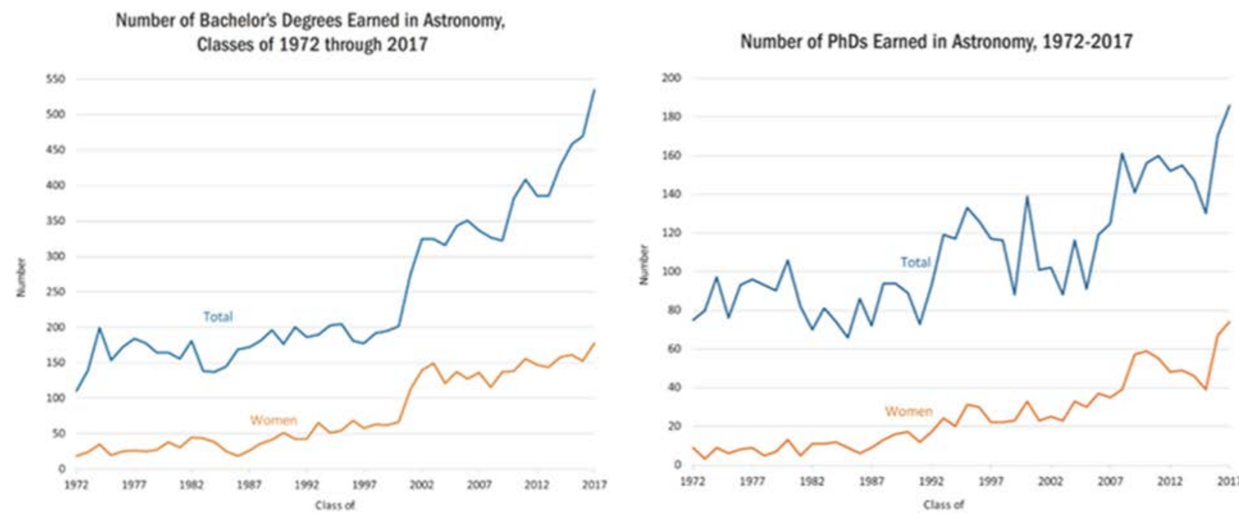


FIGURE 3.6 (*left*) Number of bachelor's degrees earned in Astronomy from 1972-2017. (*right*) Number of doctorates earned in Astronomy from 1972-2017. SOURCE: *Left:* Courtesy of Porter, A., and Ivie, R., 2019, "Women in Physics and Astronomy, 2019" © Statistical Research Center at the American Institute of Physics. *Right:* Courtesy of Porter, A., and Ivie, R., 2019, "Women in Physics and Astronomy, 2019" © Statistical Research Center at the American Institute of Physics.

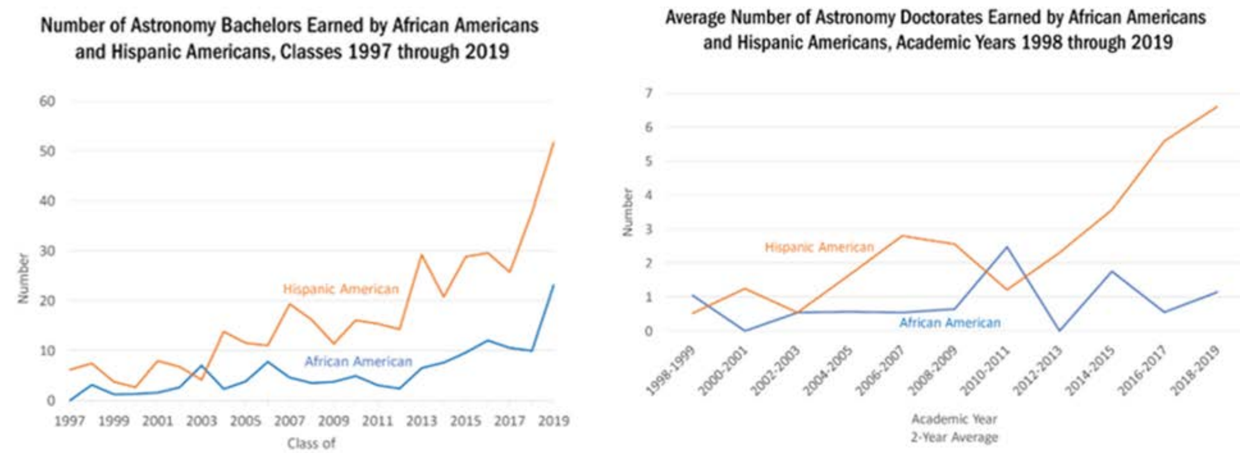


FIGURE 3.7 The numbers of astronomy degrees earned by African Americans and Hispanics. (*left*) Bachelors. (*right*) Doctorates. SOURCE: Courtesy of the Statistical Research Center at the American Institute of Physics.

A broader snapshot view of the academic pipeline into astronomy and astrophysics (see Table 3.2 for statistics and sources) reveals important patterns of ongoing disparities in the profession. Only about 2 percent of all first-year college students in the U.S. expressed an interest to major in the physical sciences. Of these, about 11 percent of White students complete a physics or astronomy bachelor's degree, whereas only 4 percent of students from underrepresented groups with similar interests do so, a disparity of about a factor of 3. While there is no longer a significant ethnic/racial disparity between the baccalaureate and Ph.D. stages (the combination of ~30 percent graduate admission rate and ~60 percent Ph.D. completion rate for those admitted are similar for all groups (see Table 3.2 and discussion below), the very large disparity at the source (undergraduate) level nonetheless culminates in a very low number of Black, Hispanic, and Indigenous Ph.D.s in physics and astronomy, with obvious long-term consequences for the diversity of the profession at the postdoctoral level and beyond. These data signify a systemic failure to fully tap the available talent pool generally, and diverse talent in particular.

Finding: Only four percent of college freshmen who are underrepresented minorities intending to major in physical sciences complete a Physics/Astronomy degree, compared to 11 percent overall. Of those, only ~16 percent continue to a Ph.D., comparable to 18 percent for all U.S. citizens.

Conclusion: There exists an enormous opportunity to tap into the nation's diverse talent already in the higher-education pipeline.

TABLE 3.2 Physics and Astronomy Synthetic Cohort from College First Year to Ph.D.

	U.S. Citizens	White	AHN[a]
First year, first-time undergraduates, all majors (2007)[b]	2,764,690	1,655,714	766,844
Estimated number intending physical sci major (2007)[c]	66,000	41,000	11,500
… of whom, __% complete physics or astronomy degrees.	*10%*	*11%*	*4%*
Bachelor's degrees in physics and astronomy (2012)[d]	6,664	4,596	473
… of whom, __% are admitted to graduate programs	*29%*	*NA*	*NA*
Entering grad program[e] in physics or astronomy (2012)[d]	1,937	NA	NA
... of whom, __% complete the Ph.D. in 6 years	*59%*	*NA*	*NA*
Ph.D. degrees in physics and astronomy (2018)[d]	1,151	805	76
Overall retention from bachelor's to Ph.D.	*17%*	*18%*	*16%*

[a] AHN = African Americans/Blacks, Hispanic/Latinx, and American Indian/Alaska Natives/Native Hawaiians. These were the names of the categories used by NSF at the time these data were collected.
[b] Enrollment of first-time, first-year undergraduate students at all institutions, by citizenship, ethnicity, race, sex, and enrollment status, Table 2-2, 2004-14 (2013 Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017. Special Report NSF 17-310. Arlington, VA., WMPD) available at www.nsf.gov/statistics/wmpd/.
[c] Based on numbers in Appendix table 2-16 Freshmen intending S&E major by, by field, sex, and race or ethnicity, 1998-2012, [NSF Science and Engineering Indicators, 2016]. Unfortunately, the number of entering first-year students who intend to major specifically in physics or astronomy is not known.
[d] AIP Enrollments and Degrees Survey.
[e] Includes M.S. and Ph.D. students.

Table 3.2 is a snapshot representation of a cohort of American students in physics and astronomy, from entering first-year students in 2007 to Ph.D. in 2018. This is a synthetic cohort in that it does not represent a literal longitudinal tracking of the same individuals over time; rather, the experience of the cohort is inferred by comparing national demographics data at time points separated by the typical

duration of various academic stages. There are also limitations to a simple, linear "pipeline" progression model; however, it does provide a convenient basis for useful comparisons. For example, the progression depicted in Table 3.2 does not disaggregate students who begin their undergraduate education at 2-year community colleges, where roughly half of all underrepresented minority students begin their post-secondary education.[34] In addition, physics and astronomy are linked by the fact that students who eventually earn Ph.D.'s in astronomy and astrophysics often begin as physics majors. While data captured by AIP on Asian Americans who earn Ph.D.s in physics (5 percent Ph.D.s among a total population of 6 percent of Asian Americans in the overall population)[35], the numbers do not directly reveal the challenges or make-up of this uniquely diverse group, defined by the Office of Management and Budget (OMB) as "a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent" in the U.S. Census.[36] Inclusive recommendations that benefit all underrepresented groups while focusing on the extremes will allow for broad reaching benefit.

The data also show that past and current efforts to engage with local and indigenous communities have not been effective enough, specifically in the context of education and training opportunities (See Section 3.4 for more general discussion and recommendations around improved engagement). For example, astronomical first light on at Hawai'i's Maunakea Observatories—a site of great cultural significance to the Kanaka Maoli—was almost exactly 50 years ago, yet in that time Ph.D.s in astronomy or astrophysics have been awarded to a total of three Native Hawaiians,[37] one of whom is currently on the faculty of a U.S. college/university astronomy program; Native Hawaiians thus comprise ~0.05 percent of astronomy faculty, compared to ~0.2 percent of the U.S. population.[38] Indigenous people in the U.S. more generally represent ~0.25 percent of Ph.D. astronomers, compared to ~2.0 percent of the population in 2019.[39,40] In addition, engagement of Native Hawaiians and Native Americans in astronomy at the undergraduate level is among the lowest of all physical sciences, averaging ~two individuals per year. Relative to overall field size, the underrepresentation in astronomy is worse than in most other physical sciences, including chemistry, Earth sciences, and physics (Table 3.3). The importance of engagement with local and indigenous communities in the context of sites where ground-based research facilities are built and operated is discussed in Section 3.4 below, including a major recommendation for the development of a new model for respectful, collaborative decision-making in partnership with Indigenous and other local communities.

There have been efforts in the past decade to increase the economic, cultural, and educational benefits of astronomy facilities for local and Indigenous communities. Examples include the `Imiloa Center in Hilo, Hawai'i and the Indigenous Education Institute.[41,42] Another example is the program in place at the Kitt Peak National Observatory that coordinates with the Tohono O'odham Nation's tribal employment office on preferential hiring of Native Americans at Kitt Peak, as well as opportunities for technical training. The Akamai Workforce Initiative, supported in part by funding from NSF, the Keck and TMT Observatories, and others, has helped hundreds of Native Hawaiians attain employment within

[34] Stassun (2003), "CSMA to Host Special Session in Seattle on Role of Minority Serving Institutions" in AAS Spectrum newsletter, January 2003. https://aas.org/sites/default/files/2019-09/spectrum_Jan03.pdf.

[35] AIP Statistical Research. "Race and Ethnicity of Physics PhDs, Classes of 2018 and 2019 Combined." Race and Ethnicity of Physics PhDs, Classes of 2018 and 2019 Combined, American Institute of Physics (aip.org)

[36] https://www.census.gov/topics/population/race/about.html.

[37] The first astronomy Ph.D. to a Native Hawaiian was awarded nearly three decades ago; the second was awarded in 2015, becoming the first Native Hawaiian to receive a Ph.D. in astronomy from the State of Hawaii's own university system; and the third was the very next year through the NSF PAARE supported program at Vanderbilt and Fisk Universities, which that same year also awarded the first astrophysics degree to a member of the Sioux Nation. See Section 3.3.4.

[38] See report of the Panel on the State of the Profession and Societal Impacts.

[39] "The Nelson Diversity Surveys" Nelson, D. J.: http://cheminfo.chem.ou.edu/faculty/djn/diversity/top50.html.

[40] See https://www.census.gov/newsroom/facts-for-features/2019/aian-month.html.

[41] See http://www.imiloahawaii.org.

[42] See http://indigenousedu.org/.

the broader STEM workforce (Figure 3.8).[43] There are also examples from other countries, such as the ALMA observatory in the Atacama region in Chile, which involves the Likan Antai people in many of its activities, including efforts to preserve the Indigenous language and cosmic worldview.

TABLE 3.3 Bachelor Degrees Earned by Indigenous People per 1,000 Degrees in the United States

	Degrees per 1000 (2003)	Degrees per 1000 (2013)	Change
Chemistry	8.1	6.4	Loss
Physics	1.8	2.3	Gain
Earth Sciences	1.9	2.3	Gain
Atmospheric Science	0.4	0.4	No change
Ocean Sciences	0.2	0.2	No change
Astronomy	0.15	0.15	No change
Other Physical Sciences	0.2	0.4	Gain

SOURCE: AIP.

FIGURE 3.8 Participants in the Akamai Workforce Initiative in 2019. The Akamai Internship Program offers STEM college students from Hawai'i a summer work experience at an observatory, company or scientific/technical facility in Hawai'i. SOURCE: https://www.akamaihawaii.org/akamai-photo-gallery/, Courtesy of the Institute for Scientist & Engineer Educators, photo by David Harrington.

Conclusion: Fewer Native Americans are receiving baccalaureate degrees in astronomy than for any other physical science. Astronomy has not fully engaged with communities with a cultural stake in the places where astronomers build facilities. Funding to PIs at tribal colleges, from

[43] See https://www.akamaihawaii.org/.

Indigenous communities, or at institutions that predominantly serve Indigenous populations, would enable long-term research partnerships and culturally supported pathways for full participation of Indigenous people in science careers.

3.3.4 The Role of Federal Agencies and Professional Societies for Diversity and Inclusion in the Profession

There have been positive and negative trends in the diversity of the Ph.D. pipeline over the past 25 years (see Figure 3.9). While the factors driving these trends are no doubt complex, a simple comparison of the timing of recent gains and losses in the diversity of the academic pipeline at the baccalaureate and especially the Ph.D. levels suggests that at least some, and perhaps much, can be attributed to funding initiatives by NASA and NSF that, starting about 20 years ago, began to invest specifically in workforce diversity, and in particular through partnerships with minority-serving institutions (MSIs).[44] Taking one of the first of these programs as a specific example, the Fisk-Vanderbilt Bridge Program began in 2004 with NASA support until 2007 and subsequent NSF support with a final award in 2013.[45] The program's first cohort began to complete their Ph.D.s in 2009, and by 2015 the program was responsible for an average of six Ph.D.s per year to underrepresented minority students, by itself representing an increase of ~30 percent over the number that was being awarded nationally when the program began. By that time, the cumulative impact of additional programs (see below) was becoming evident (see Figure 3.9).

These programs served to engage underrepresented students in research experiences while enhancing the astronomy and astrophysics research capacity of the MSIs. The choice to specifically form partnerships with MSIs was in recognition of the outsized role that these institutions play in the recruitment, support, and preparation of underrepresented minorities for science and engineering careers. For example, all 10 of the top 10 producers of African American baccalaureate degrees in physics are Historically Black Colleges and Universities (HBCUs).[46]

Finding: Minority Serving Institutions—including Historically Black Colleges and Universities, Hispanic Serving Institutions, and Tribal Colleges and Universities—are a large and diverse talent pool for the field. For example, all 10 of the top 10 producers of African American baccalaureates in physics are HBCUs.

Importantly, these funding initiatives were operated at the relevant division levels of the agencies with purview over astronomy and astrophysics, not centralized at the top agency levels where their impact specifically on the astronomy and astrophysics workforce at the undergraduate and graduate levels might be diffused. The NASA Science Mission Directorate (SMD) program was called MUCERPI (Minority University and College Education and Research Partnership Initiative),[47] and the NSF AST program was called PAARE (Partnerships for Astronomy and Astrophysics Research and Education).[48] Although not fully comparable to PAARE or MUCERPI, the DOE Office of Science does run a Visiting Faculty Program (VFP, formerly Faculty and Student Teams [FaST]) that supports individual MSI faculty or small faculty-student teams.[49] Some of the most well-known programs of the past 20 years—such as the

---

[44] MSIs include Historically Black Colleges and Universities (HBCUs), Hispanic Serving Institutions (HSIs), and Tribal Colleges and Universities (TCUs).

[45] Stassun (2017) https://pubs.acs.org/doi/full/10.1021/bk-2017-1248.ch006

[46] K.G. Stassun, Congressional Testimony, 16 March 2010. http://astro.phy.vanderbilt.edu/~stassuk/KGStassun_CongressionalTestimony_30Jul2010_revised.pdf

[47] Sakimoto, P. J. and Rosenthal J. D., 2005, Physics Today, September, p. 49-53.

[48] See https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501046.

[49] See https://science.osti.gov/wdts/vfp.

Fisk-Vanderbilt Masters-to-Ph.D. Bridge Program (see above),[50] the Columbia Bridge to the Doctorate Program,[51] the CalBridge Program,[52] and others—have been "bridge" type programs through which underrepresented students at the undergraduate level are trained and supported specifically across the transition into graduate-level training (Figure 3.10). As noted above (see also Table 3.2), ethnic/racial disparities from the baccalaureate to the Ph.D. stages of education and training in physics and astronomy are no longer significant, which is a significant accomplishment in itself. All of these programs got their funding start through MUCERPI, PAARE, FaST, or some combination of these and institutional resources.

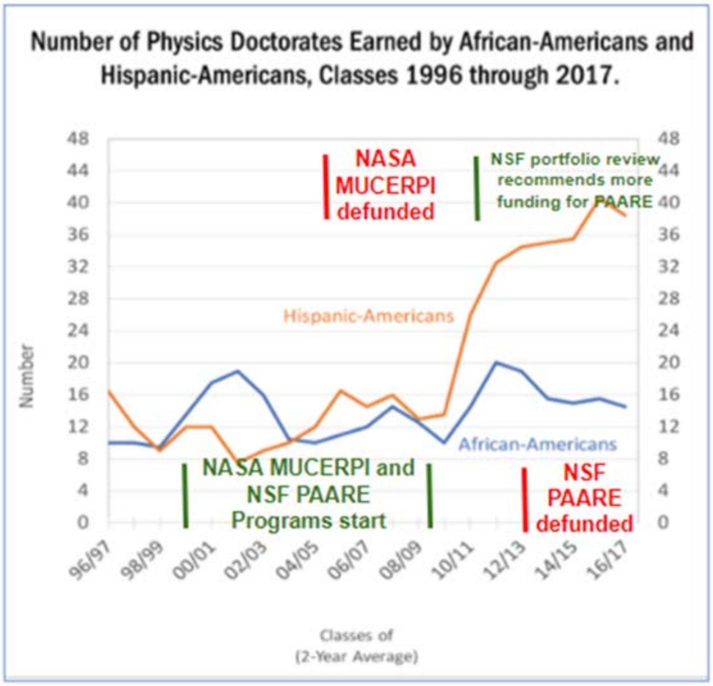


FIGURE 3.9 The numbers of doctoral degrees earned by African-Americans and Hispanic-Americans, with key dates pertaining to NASA and NSF funding programs dedicated to workforce diversity. Note that, due to the amount of time required for any individual to complete a Ph.D. and for any one program to reach steady state, the rise/decline in the number of Ph.D.s earned lags the initiation/termination of programmatic interventions by 5-10 years. SOURCE: https://www.aip.org/statistics/reports/trends-physics-phds-171819. Adapted from the Statistical Research Center at the American Institute of Physics.

[50] Stassun, K.G. 2011 American Journal of Physics 79, 374.
[51] See https://bridgetophd.facultydiversity.columbia.edu/.
[52] See https://physicstoday.scitation.org/doi/10.1063/PT.3.4319.

FIGURE 3.10 Training programs that provide a "bridge" for students from undergraduate to doctoral studies—such as the Fisk-Vanderbilt Masters-to-Ph.D. Bridge Program (*upper left*), the CAMPARE program in California (*upper right*), and the Columbia Bridge to the Doctorate Program (*bottom*)—have emerged over the past two decades as a promising mechanism for advancing inclusive excellence in astronomy and astrophysics Ph.D. programs. SOURCE: https://www.nsf.gov/mps/ast/broadening_participation/index.jsp. Courtesy of Donald Pickert/Vanderbilt University; Cal-Bridge Summer (CAMPARE) Program; Columbia University.

In recognition of the early successes of these programs,[53] the American Physical Society (APS) launched a program to emulate these efforts and incentivize similar programs in physics departments nationally.[54] NSF AST's 2013 portfolio review specifically recommended line-item funding for "workforce diversity" as part of its broader recommendation for augmenting the small+midscale budget for NSF AST.[55] Unfortunately, all of these division-level workforce diversity funding programs have since been defunded, as a result of budget pressures, top-level agency programmatic consolidation, or both. Reinvesting in these programs could yet yield significant benefits for the diversity of the field. As noted above, the amount of time required for individuals to complete Ph.D. training and for programs to ramp up implies that programs likely need to be supported for at least 5-10 years to enable reaching steady state impact.

[53] Rudolph, A., Holley-Bockelman, K., Posselt, J. 2019. Nature Astronomy, 3, 1080. https://www.nature.com/articles/s41550-019-0962-1.

[54] See https://physicstoday.scitation.org/doi/10.1063/PT.3.3464.

[55] See https://www.nsf.gov/mps/ast/ast_portfolio_review.jsp.

Finding: Previous NASA, NSF, and DOE funding programs (e.g., NSF PAARE, NASA MUCERPI, DOE FaST) focused on training in state-of-the-art research methods and preparation for future leadership in research including computation, instrumentation, etc., especially through partnerships with minority-serving institutions, have been defunded. Importantly, these funding initiatives were not centralized at the top agency levels where their impact specifically on the astronomy and astrophysics workforce at the undergraduate and graduate levels might be diffused; rather, they were initiated and operated at the relevant division levels of the agencies with purview over astronomy and astrophysics.

Recommendation: NASA, NSF, and DOE should reinvest in professional workforce diversity programs at the division/directorate levels with purview over astronomy and astrophysics. Because academic pipeline transitions are loss points in general, supporting the creation and continued operation of "bridge" type programs across junctures in the higher-education pipeline and into the professional ranks appear especially promising.

One outcome of efforts to accelerate the participation of underrepresented groups in graduate education is that many departments have modified their graduate program application requirements to more effectively attract talented, high-achieving students from an increasingly diverse pool of candidates. Indeed, there is an emerging sensibility around the imperative of equity-based holistic review—a practice that has applicability not only for admissions, but also hiring, awards, grants, and leadership positions. The AAS task force on diversity and inclusion in graduate education compiled lessons learned from the movement to improve graduate admissions, recruitment, and mentoring, as well as program climate and data use. Their recommendations were endorsed by the AAS in January 2019,[56] amplifying recommendations and toolkits from the first Inclusive Astronomy meeting that was convened in 2015 and endorsed by the AAS (Figure 3.11).[57,58] Importantly, one core recommendation from Inclusive Astronomy in the "Power, Policy, and Leadership" category was that the Astro2020 decadal survey should "include recommendations (i.e., not merely findings as in previous decadal surveys)." More generally, the Inclusive Astronomy recommendations included a roadmap for establishing a "community of inclusive practice," engaging the astronomy community as a whole (including AAS committees such as SGMA, CSMA, CSWA, and WGAD, among others)[59] in ongoing two-way engagement between professional societies and the members that comprise them to create a much more powerful voice for the decadal goals, as well as create a more engaged, diverse, and inclusive community of scientists working toward common purposes.

Finding: Leadership by the astronomy community in the past decade has produced exemplary efforts for inclusive excellence in graduate education, including the promotion and implementation of equity-based holistic review practices for admission, evidence-based practices for mentoring, and data-driven approaches to improved program climate.

[56] See https://aas.org/sites/default/files/2019-09/aas_diversity_inclusion_tf_final_report_baas.pdf.
[57] See https://tiki.aas.org/tiki-index.php?page=Inclusive_Astronomy_The_Nashville_Recommendations.
[58] See https://aas.org/posts/news/2017/02/inclusive-astronomy-nashville-recommendations.
[59] Committee for Sexual-Orientation & Gender Minorities in Astronomy (SGMA), Committee on the Status of Minorities in Astronomy (CSMA), Committee on the Status of Women in Astronomy (CSWA), Working Group on Accessibility and Disability (WGAD). https://aas.org/committees-and-working-groups.

FIGURE 3.11 Following the example of three Women in Astronomy meetings over the past 20 years, the first Inclusive Astronomy meeting, held at Vanderbilt University in 2015, produced the "Nashville Recommendations" for making the astronomy and astrophysics community more diverse and inclusive from the undergraduate to leadership levels. SOURCE: https://www.planetary.org/articles/0625-inclusive-astronomy; Courtesy of Donald Pickert/Vanderbilt University.

In addition, an important principle to emerge from multiple National Academies reports addressing discrimination and harassment (see below and Box 3.1), is that early-career scientists from undergraduates to graduate students to postdocs need greater access than is currently the norm for funding support that provides independence and flexibility (so as to lessen over-reliance on individual advisors and/or hierarchical training relationships), while at the same time increasing access to more structured opportunities for mentoring networks, evidence-based pedagogy, training for different career paths, etc. Exemplar approaches suggested by, e.g., the National Academies report on the *Science of Effective Mentoring* (2018) and the AIP National Task Force to Elevate African American Representation in Undergraduate Physics & Astronomy (TEAM-UP) report (2019) includes connecting students to structured cohort-based research training programs, such as the National Institutes of Health (NIH) Maximizing Access to Research Careers (MARC) awards (undergraduate) and "T" training grant programs (graduate), as well as independent fellowship funding at the postdoctoral level, and ensuring that such funding is awarded to a broadly diverse set of institutions to ensure equitable access.[60]

> **Recommendation: NSF, NASA, and DOE should implement undergraduate and graduate "traineeship" funding, akin to the NIH MARC and NIH "T" training grant programs, to incentivize department/institution-level commitment to professional workforce development, and prioritize interdisciplinary training, diversity, and preparation for a variety of career outcomes.**

> **Recommendation: NASA and NSF should continue and increase support for postdoctoral fellowships that provide independence while encouraging development of scientific leaders who advance diversity and inclusive excellence (e.g., NASA Hubble Fellows program, NSF Astronomy and Astrophysics Postdoc program).**

[60] For example, the NASA Hubble Fellows Program (NHFP) is conducting an independent, outside review of its policies and procedures to improve equitable access and diversity in fellowship recipients and host institutions.

3.3.5 Addressing Racism, Bias, Harassment, and Discrimination

No discussion of the factors shaping the profession would be complete without addressing the uncomfortable but all-too-real challenges of racism, bias, harassment, and discrimination in the field. Building toward a fully diverse and inclusive workforce is unequivocally a long-term priority for the profession, and the persistence of discrimination in the field in any form—including in the forms of racism, bias, and harassment—will continue to fundamentally hamper progress toward that important goal. As noted by the SoPSI report, "discrimination in the profession (be it structural or between individuals, overt or implicit) impacts (i) professional well-being by producing stress and other negative health outcomes; (ii) equitable participation and advancement by not accounting for these differences in experience and mental/emotional load when evaluating performance and outcomes; and (iii) economic prosperity and innovation by limiting the degree to which minoritized populations can obtain and maintain jobs in the profession and further a deeper understanding of the universe."

These challenges extend beyond astronomy and have been addressed in numerous reports over the past five years, including several National Academies studies highlighted in the Box 3.1, the 2019 report of the AAS Task Force on Diversity and Inclusion in Graduate Education,[61] and an extensive report from the AIP's TEAM-UP (see Figure 3.12). The report's recommendations[62] are grouped into five key "factors" that include a sense of belonging, physics identity, academic support, personal support, and leadership and structures. The principles here can also be applied to diversity efforts beyond the undergraduate experience, including staff hiring such as engineers, administrators, and those from other scientific backgrounds as well. The TEAM-UP report is an especially important and timely one for the field, at a time when growing awareness of the effects of systemic racism continue to have significant, substantive, and negative effects on the African American community specifically and communities of color generally, and how these societal and structural problems present real barriers to inclusion for the physics and astronomy community in particular.

Progress is also being made in implementing many of the recommendations from these various reports. At the STEM-wide level the American Association for the Advancement of Science (AAAS) STEM Equity Achievement (SEA) Change program[63] is a comprehensive initiative aimed at advancing inclusion and persistence of scientists from historically underrepresented groups, and incorporates proven self-assessment elements to establish goals and measure progress towards reaching them. In physics the APS has initiated an Effective Practices for Physics Programs (EP3) program[64] that is aimed to implement many of the TEAM-UP recommendations (including physics and astronomy departments). These examples serve as models for the types of follow-up activities needed within astronomy itself.

Up to now this discussion has mainly focused on diversity and inclusion efforts in professional education and academic departments, but improvements are needed in the research sector as well. The funding agencies have also taken some proactive steps to mitigate bias in the awarding of resources for research. Proposals for observations with NASA's Hubble Space Telescope were the first to employ a dual anonymous proposal review process in 2018, after analysis of gender-based proposal successes over 10 years demonstrated a small but consistent pattern of male PI success exceeding that of women's success.[65] The effect on women's success rates after implementing the dual anonymous review process varies with proposal category and observing cycle. One large and noticeable effect of the implementation of proposal review focused on the science and not on the scientists was a large increase in the percentages of new principal investigators on a mature facility (Figure 3.13). NASA SMD is following with a trial implementation of dual-anonymous proposal review procedures for selected programs in astrophysics and beyond, and some NSF supported observatories are following suit as well. As a respected field influential

[61] See https://aas.org/sites/default/files/2019-09/aas_diversity_inclusion_tf_final_report_baas.pd.
[62] See https://www.aip.org/diversity-initiatives/team-up-task-force.
[63] See https://seachange.aaas.org/.
[64] See https://ep3guide.org/.
[65] Reid, I. N. et al. 2014 PASP 126, 923 https://ui.adsabs.harvard.edu/abs/2014PASP..126..923R/abstract.

in public opinion, astronomy's move toward such equitable and inclusive practices may influence other professions. It is encouraging to see NASA and NSF piloting and assessing the impact of this approach.

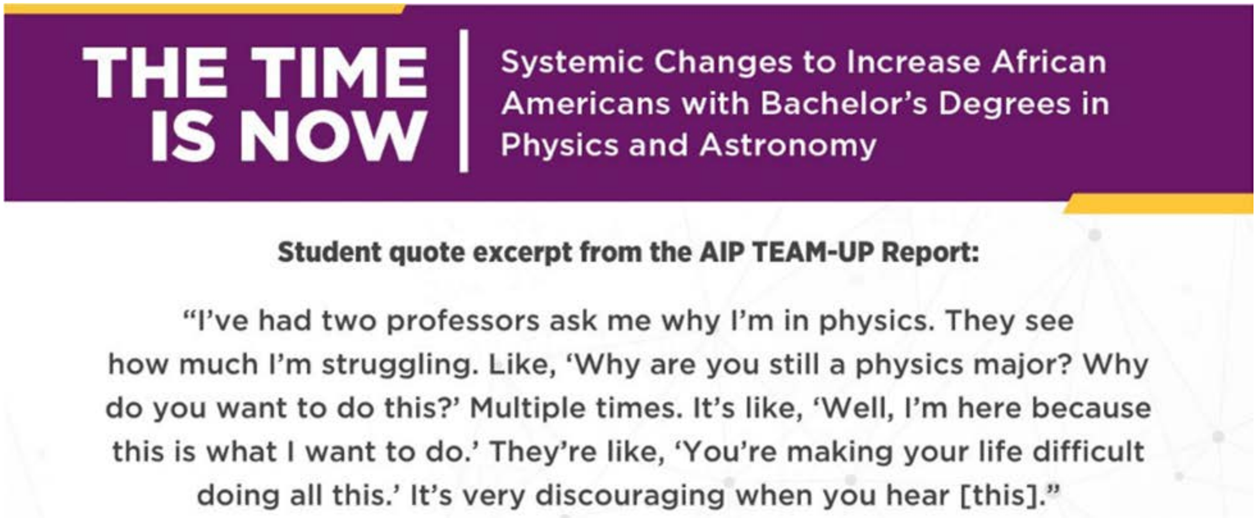


FIGURE 3.12 AIP's TEAM-UP (Task Force to Elevate African American Representation in Undergraduate Physics & Astronomy) report explores the ongoing effects of racism in society and in physics and astronomy. The report offers concrete recommendations to make the physics and astronomy community more inclusive and to increase the representation of African Americans in the field. SOURCE: Courtesy of the American Institute of Physics. bit.ly/TEAMUPReport.

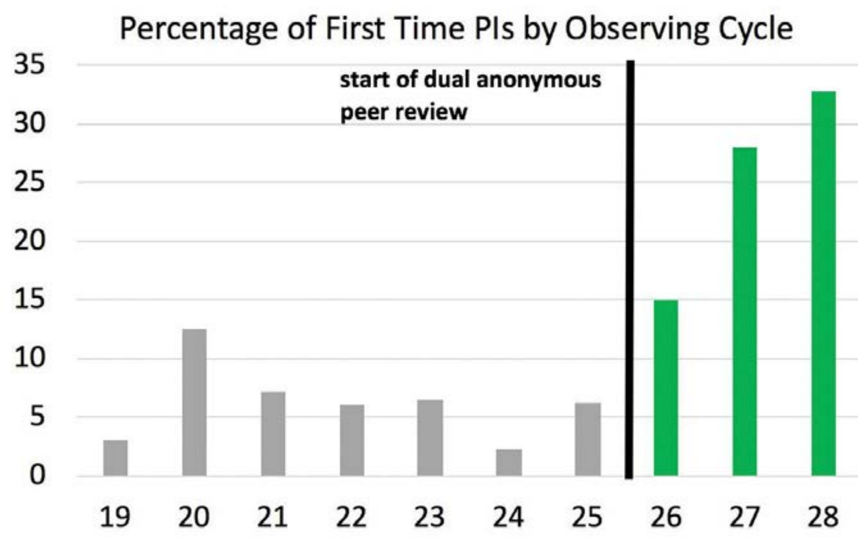


FIGURE 3.13 Percentage of first time principal investigators with successful proposals to use the Hubble Space Telescope, as a function of observing cycle. Observing cycles 26, 27, and 28 utilized a dual-anonymous peer review process, which has had a significant impact on bringing in new investigators, even two and a half decades after the launch of the observatory. SOURCE: https://www.stsci.edu/contents/newsletters/2020-volume-37-issue-02/hst-stsci-update?Volume=7c634b8f-a80c-496b-b5c7-5aaee88610f4&filterUUID=7b401d2c-07c2-4980-b769-77bc6ebf33ae&filterPage=newsletters&filterName=filter-articles; Courtesy of R.A. Osten/STScI.

Finding: NASA and some NSF supported observatories have implemented a trial of dual-anonymous procedures as part of its proposal merit review process, in a proactive effort to mitigate bias in proposal evaluation and selection.

Even so, much more remains to be done. As indicated by a number of widely reported cases in astronomy and astrophysics in the past decade, the astronomy and astrophysics profession cannot yet claim to have eliminated the scourges of sexual harassment and discrimination that continue to afflict many professions. Indeed, as powerfully illustrated in the recent National Academies (2018) report on *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*, the "obvious" or most blatant cases often represent only the tip of the proverbial iceberg, and the data reveal that experiences of sexual harassment and discrimination remain much more widespread than many scientists imagine or would like to admit.[66] For example, the (2018) report reveals that, "the academic workplace (i.e., employees of academic institutions) has the second highest rate of sexual harassment at 58 percent (the military has the highest rate at 69 percent)." Academic science is, evidently, a high-risk workplace for a certain type of occupational safety hazard.

To be sure, the situation today is certainly better in many ways compared to a time when harassment was more pervasive and blatant, and some types of discrimination were even legally permitted. But such a comparison is small comfort given how prevalent harassment and discrimination remain, and it certainly does not represent a high bar for fairness, let alone excellence.

Conclusion: The persistence of harassment and discrimination in astronomy and astrophysics is intolerable, and must not be tolerated if the astronomy and astrophysics profession is to retain and successfully draw from the full diversity of talent available, not to mention avoiding the toxic and corrosive effects that such behaviors have on individuals, organizations, and the entire profession.

What needs to be done to fully address this issue once and for all is not a mystery; there are well-established best practices, documented solutions, and veritable how-to guides that can be implemented at the individual, organizational, and profession-wide levels (see Box 3.1) that the astronomy and astrophysics community could endorse, adopt, and most importantly, work deliberately to implement. These include an especially important role for the federal funding agencies that, backed by existing federal laws, can use the power of the purse as a forcing function to help drive needed change.

Finding: There are best practices to eradicate and prevent harassment and discrimination, and to promote healthy and inclusive work cultures across the astronomy and astrophysics profession, described in detail in previous National Academies and other reports.

That the solutions sit before us, yet harassment and discrimination persist, is a disgrace. That perpetrators of harassment and discrimination are not decisively and consistently stopped—indeed, that they are sometimes tolerated or even professionally rewarded despite their shameful behavior—is a profound injustice to people who have been harmed and is morally wrong. And it is ultimately, as multiple recent reports argue, a failure of leadership to muster the courage to break free of organizational blame-avoidance: "Too often, interpretation of Title IX and Title VII has incentivized institutions to create policies and training on sexual harassment that focus on symbolic compliance with current law and avoiding liability, and not on preventing sexual harassment" (see Box 3.1). Just as hazardous workplaces such as factories and construction sites carefully track and publicly report the number of work-days without injuries, let astronomy strive as a profession for nothing less than a 100 percent safety record (i.e., no tolerance for those who abuse their position and their colleagues) with regards to harassment and

[66] See the widely circulated iceberg infographic from National Academies report Sexual Harassment of Women: Climate Culture, and Consequences in Academic Sciences, Engineering, and Medicine (2018). https://www.nap.edu/visualizations/sexual-harassment-iceberg/.

discrimination in our classrooms, laboratories, observatories, research centers, and everywhere that members of the profession—and those who aspire to it—do their work.

Recommendation: NASA, NSF, DOE, and professional societies should ensure that their scientific integrity policies address harassment and discrimination by individuals as forms of research/scientific misconduct.

BOX 3.1 Harassment and Discrimination

Since 2018, the National Academies of Sciences, Engineering, and Medicine have released multiple consensus reports that have taken a systemic look at addressing harassment and discrimination as key issues in higher education and academic research: *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*; *Graduate STEM Education for the 21st Century*; *The Science of Effective Mentorship in STEMM*; and *Minority Serving Institutions: America's Underutilized Resource for Strengthening the STEM Workforce*; as well as the *Exoplanet Science Strategy* report.[1]

A common theme in these reports is to situate the issue of sexual harassment and discrimination within the broader cultures of academia and scientific work environments. As described in the National Academies report on *Sexual Harassment of Women*: "Four aspects of the science, engineering, and medicine academic workplace tend to silence targets of harassment as well as limit career opportunities for both targets and bystanders: (1) the dependence on advisors and mentors for career advancement; (2) the system of meritocracy that does not account for the declines in productivity and morale as a result of sexual harassment; (3) the 'macho' culture in some fields; and (4) the informal communications network, through which rumors and accusations are spread within and across specialized programs and fields."[2]

The reports furthermore identify the incentive and reward systems of academic science as critical drivers of individual and organizational behavior. Five features in particular, especially in combination, are found to be most predictive of toxic workplace environments: (1) a perceived tolerance for harassment or discrimination, which is the most potent predictor of these occurring in an organization; (2) male-dominated work settings in which men are in positions of authority—as deans, department chairs, principal investigators, and dissertation advisors—and women are in subordinate positions as early-career faculty, graduate students, and postdocs; (3) environments in which the power structure of an organization is hierarchical with strong dependencies on those at higher levels; (4) a focus on "symbolic compliance" with federal laws that should have teeth if properly implemented—especially Title IX and Title VII—resulting in policies and procedures that protect the liability of the institution but are not effective in preventing harassment and discrimination; and (5) leadership that lacks the intentionality and focus to take the bold and aggressive measures needed to reduce and eliminate harassment and discrimination.

In addition, the reports reach broad consensus that the federal legal framework alone is essential, but is by itself not adequate for reducing or preventing sexual harassment and discrimination. Indeed, one of the major recommendations of the 2018 report is that "academic institutions, research and training sites, and federal agencies should move beyond interventions or policies that represent basic legal compliance and that rely solely on formal reports made by targets."[3] The report argues that there must be proactive, not reactive, efforts to create "diverse, inclusive, and respectful work environments," to improve transparency and accountability through appropriate statistical reporting regarding incidence numbers and rates, to diffuse hierarchical and dependent relationships, to provide support for victims and targets, and importantly to strive for strong and diverse leadership.

These reports further highlight the role that federal agencies, which control research funding, can play in enacting long-lasting change. Indeed, some significant steps are already underway, at least with respect to holding grantee institutions more accountable. Finally, discussions in the literature and in the

community have in recent years begun to broaden the view of harassment and discrimination to be more inclusive of all identity categories, toward the more all-encompassing notion of "identity-based discrimination" (see, e.g., National Academies workshop report on *The Impacts of Racism and Bias on Black People Pursuing Careers in Science, Engineering, and Medicine: Proceedings of a Workshop*, 2020[4]). Identity-based discrimination includes both differential treatment (including harassment) on the basis of identities and ostensibly neutral practices that produce differential impacts owing to identity (see, e.g., National Academies *Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors*, 2020[5]). More such studies are needed, as well as a more complete development of identity-based discrimination within the larger legal and regulatory framework. As noted by the SoPSI report, "cultural shifts around identity-based harassment require second-order theories of change (i.e., addressing underlying priorities and norms, not just reforming policy and practice) and an intersectional lens (i.e., attending to experiences of people with multiple marginalized identities)."

[1] National Academies of Sciences, Engineering, and Medicine (NASEM). 2018. *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24994. NASEM. 2018. *Graduate STEM Education for the 21st Century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25038. NASEM. 2019. *The Science of Effective Mentorship in STEMM*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25568. NASEM. 2019. *Minority Serving Institutions: America's Underutilized Resource for Strengthening the STEM Workforce*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25257. NASEM. 2018. *Exoplanet Science Strategy*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25187.
[2] NASEM. 2018. *Sexual Harassment of* Women, p. 3.
[3] Ibid., p. 181.
[4] NASEM. 2020. The Impacts of Racism and Bias on Black People Pursuing Careers in Science, Engineering, and Medicine: Proceedings of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/25849.
[5] NASEM. 2020. Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors. Washington, DC: The National Academies Press. https://doi.org/10.17226/25585.

3.3.6 Demographics Data, Outcomes, and Accountability

Across astronomy and astrophysics, there is a growing emphasis on making the field a place where everyone can thrive. However, while ideas abound for improving inclusion and access, it is not possible to assess whether any strategy is working without the associated data to measure what is happening. Obtaining these critically needed data remains a challenge. For example, the SoPSI panel requested data on astronomy-related programs from NASA, NSF, and DOE as well as management organizations for major astronomical facilities. Requested data included demographics of staff, contractors, review panels, proposers, and awardees of grants and fellowships, along with data on agency programs and funding that promote broader access to opportunities and reduce barriers to achieving success in the field for underrepresented groups. Unfortunately, the data produced by the federal agencies were minimal.

While all three agencies collect some demographic data (usually binary gender, race, and ethnicity) on staff and applicants for funding, several issues are clear. First, the agencies do not collect and track the same quantity or categories of demographic data. NSF has gathered demographic information for many years but only publishes it in aggregated form.[67,68] In response to a 2015 critique by

[67] NSF's National Center for Science Engineering Statistics www.nsf.gov/statistics/about-ncses.cfm#service
[68] Report on Merit Review, 2019 www.nsf.gov/statistics/about-ncses.cfm#service.

the Government Accountability Office,[69] NASA began collecting additional demographic data through its proposal submission website, the NASA Solicitation and Proposal Integrated Review and Evaluation System (NSPIRES),[70] but the data are not yet publicly available. The DOE Portfolio Analysis and Management System (PAMS) collects demographic data on applicants,[71] but is not designed for data analysis, and separate program offices within the Office of Science maintain their own databases. Individual laboratories within the Office of Science do collect and report demographic data on employees, though not on facility users (e.g., Argonne National Laboratory[72]).

Second, the policies of the agencies differ concerning public release of the information. NASA shared some information on the inferred binary gender of awardees (based on given names). By contrast, NSF declined to share specific information of this type, reserving the specific data it gathers for use in internal reviews and assessments. Third, even when the requested data was collected, it was not made readily available or the committee would have had to aggregate the information itself. Finally, none of the agencies appear to track programs and funding aimed at promoting diversity and inclusion.

There is an excellent precedent from the NIH, which has for decades collected demographic information from researchers in its external grants program (currently about 80,000 applications/year, larger than NASA's, NSF's, and DOE's grants programs combined). This process is managed by the Office of Extramural Research through their electronics grant system, the Electronic Research Administration (eRA). The funding agencies for astronomy and astrophysics are in the best position to collect, evaluate, and make available demographic data to provide a comprehensive picture of the workplaces of the Profession and the experiences of its people in it, and track funding specifically aimed at promoting community values. The NIH provides an example for the agencies to emulate.

> **Recommendation: NASA, NSF, and DOE should implement a cross-agency committee or working group tasked with establishing a consistent format and policy for regularly collecting, evaluating, and publicly reporting demographic data and indicators pertaining at a minimum to outcomes of proposal competitions.**

For any system of accountability to be meaningful, there must be clear expectations and guidelines that specify the basis for evaluation. To the extent that the profession expects improved outcomes with regards to workforce development, equity, diversity, and inclusive excellence, there must be alignment between these values and the criteria by which success and excellence are measured and evaluated. As the saying goes, "measure what you value, instead of valuing only what you happen to measure." This is especially important in the context of funding awards, since these are arguably the most effective levers for communicating expectations and for incentivizing the outcomes that the community values. NSF currently incorporates proposal evaluation criteria for outcomes related to workforce development, training, diversity, etc., in the form of its "broader impacts criterion", which explicitly values "broadening participation of underrepresented groups", among other criteria. NASA and DOE do not in general include similar evaluation criteria for funding awards at either the individual investigator or mission levels. Such criteria need to be adapted to the scale of the projects; expectations for documenting diversity, training, and workforce development efforts for a NASA Explorer or Probe project, for example, would clearly be greater than for an individual investigator grant. However, this need for flexibility does not disqualify agencies from establishing such guidelines. The SoPSI Panel report provides examples of criteria that might be established, for example describing plans for achieving diversity; participation in agency-sponsored demographic and climate assessments; mentoring and advising plans for project students and postdocs, and others. In this context we interpret diversity to

[69] "Women in STEM Research: Better Data and Information Sharing Could Improve Oversight of Federal Grant-making and Title IX Compliance", https://www.gao.gov/products/GAO-16-14.

[70] See nspires.nasaprs.com/external/.

[71] See www.energy.gov/science/office-science-funding/sc-portfolio-analysis-and-management-system-pams.

[72] See www.anl.gov/hr/argonne-employee-demographics.

encompass not only demographic diversity but also possibly institutional and/or geographic diversity, depending again on the appropriateness of such criteria for the nature and scale of the projects being proposed. For small and individual investigator projects, approaches similar to the NSF "broader impacts" requirements may be more appropriate, but the same principles of commitment to and accountability for addressing diversity and inclusion apply.

Recommendation: NASA, DOE, and NSF should consider including diversity—of project teams and participants—in the evaluation of funding awards to individual investigators, project and mission teams, and third-party organizations that manage facilities. Approaches would be agency specific, and appropriate to the scale of the projects.

Agencies may need to provide resources, including access to appropriate experts, to support the community in responding positively and successfully to enhanced criteria and accountability mechanisms, especially for proposers working at institutions where such support is not provided locally.

3.4 ASTRONOMY'S SUSTAINABLE FUTURE: CLIMATE, LIGHT, LAND AND COMMUNITIES

Astronomical activities do not occur in vacuum, disconnected from other global concerns. To the contrary, how and where astronomers conduct their work can both endanger, and be endangered by, the rights and activities and concerns of others. Indeed, some of these concerns rank among the most pressing global challenges of our time, from climate change to human rights. Consequently, the future of astronomy, like the future of so much of the world to which it is bound, will depend on the development and implementation of more sustainable practices and partnerships with the global community, commercial ventures, and Earth.

For example, astronomical data collection from the ground suffers from increasing levels of electromagnetic encroachment (e.g. "light pollution") by telecommunications and navigation systems, systems that otherwise represent high-value commercial interests and that are highly valued by billions of people around the world. And like all people, astronomers, in their individual and collective choices and actions, contribute to the carbon footprint that literally imperils life as we know it. At the same time, astronomical facilities on the ground are constructed on lands that are in some cases regarded as hallowed or revered with human, cultural significance by local and/or indigenous peoples. A renewed focus on sustainability is therefore also intertwined with the need for the development of a new model for respectful, collaborative decision-making in partnership with Indigenous and other local communities.

3.4.1 Engagement with Local and Indigenous Communities: The Model of Community Based Science

Much of astronomy is conducted on the ground—ground that is governed by laws, regulated by governmental entities, and in many cases regarded as hallowed or revered with cultural significance.[73] Nowhere do these overlapping concerns manifest themselves more poignantly and pointedly than in the case of lands that have significance for Indigenous communities. Engaging with Indigenous communities requires deliberate, respectful efforts to consider the many, complex factors both intrinsic and extrinsic to

[73] The Survey drew heavily for this section from a set of white papers submitted by the community, including especially those entitled "Kū Kia'i Mauna: Historical and Ongoing Resistance to Industrial Astronomy Development on Mauna Kea, Hawai'i", "Impacts of Astronomy on Indigenous Customary and Traditional Practices As Evident at Mauna Kea", "A collective insight into the cultural and academic journeys of Native Hawaiians while pursuing careers in physics and astronomy", "Collaboration with Integrity: Indigenous Knowledge in 21st Century Astronomy", and others cited in the SoPSI panel report.

astronomy, legal and extralegal, as well as societal histories, that span decades or even centuries. The specific case of Maunakea is an example that recently has involved tensions and has a long history wrapped up in the formation, history, and future plans for the Mauna Kea Science Reserve. See Box 3.2 "Mauna Kea Science Reserve."[74]

At the same time, strides have been taken within other scientific disciplines to create even broader "community based" models of active, up-front, and sustained engagement with local and Indigenous communities (See Figure 3.14). While there have been efforts in the past decade to increase the economic, cultural, and educational benefits of astronomy facilities for local and Indigenous communities (See Section 3.3.3), astronomy is not the only scientific discipline to have found itself at odds with the values and needs of local communities impacted by research activities. In the ways that astronomy and astrophysics research often involves literally "breaking ground" on sacred land and involves paradigms for authoritative knowledge that may differ from those of Indigenous cultures, astronomy is in many ways similar to the field of archaeology. Archaeology has evolved over time from a harmful past toward professional norms and ethical practices that are more respectful of local cultures, more reflective of the needs of local people, and more empowering of communities.[75] See Box 3.3 "The Model of Community Based Science."

> **Finding:** There have been strides within other scientific disciplines to create "community based" models of active, up-front, and sustained engagement with local and Indigenous communities based on partnership.

BOX 3.2 50 Years of Astronomy on Maunakea: the Mauna Kea Science Reserve[76]

Maunakea is part of land that was taken from the Hawaiian Kingdom during its colonization in the late 1800s. These lands were converted to the status of "public lands" when Hawaii became a U.S. state in 1959. Maunakea has great cultural and religious significance for Native Hawaiians (Kanaka Maoli); many view the development of astronomical observatories on Maunakea to be part of a larger threat to their cultural heritage. Some are concerned about environmental impacts of large facilities on the site, as well as other issues.

For the past 50+ years, the summit of Maunakea has provided conditions for astronomical observations that are nearly unrivalled compared to any other site on Earth's surface, with exceptionally dark skies and dry air, median seeing of 0.65 arcsec at 0.6 micron, and good photometric conditions about 70 percent of the time.[77] As a result, this site has attracted enormous investments in astronomical facilities by multiple countries, including the U.S. These include some of the largest and most powerful ground based visible-light and near-infrared telescopes on Earth, as well as facilities that observe at submillimeter wavelengths. Furthermore, the topography of the smooth volcano combined with the prevailing wind flows lead to the atmospheric turbulence being largely confined within about 100 m of the surface, making this site one of the most promising for future ambitious telescope projects that will utilize ground layer adaptive optics, such as the Thirty Meter Telescope (TMT).

[74] "Maunakea" is the proper name of the mountain. "Mauna Kea" is used in published or legal documents, such as in "Mauna Kea Science Reserve." http://www.malamamaunakea.org/articles/9/Maunakea.
[75] Acknowledging that the term "empower" could imply an imbalance in which one group has the right to grant power to another; the intent here is to recognize the power and autonomy that is a right of all people.
[76] See Appendix N, Section N.6.7.1 for further discussion of tensions regarding current construction on Maunakea.
[77] D. Simons, et al., The Future of Mauna Kea Astronomy, white paper.

The combined scientific impact of the Maunakea Observatories is world leading.[78] Moreover, the collective capital value of these facilities, considering construction costs alone, is in excess of $1 billion, and much more than that considering upgrades and instrument suites; this essentially represents a lower limit to the replacement cost would it become necessary to “move” all current operations elsewhere. Decommissioning these facilities is also nontrivial—the estimated cost of site restoration can be as much as $10 million or more per facility.

Ongoing improvements of capabilities of existing platforms on Maunakea, and the potential addition of TMT as a major new platform, are expected to revolutionize ground-based astronomy. In addition, activities on Maunakea are major sources of revenue for the State of Hawaii.[79] Furthermore, access to the site by the State of Hawaii’s university system is a significant opportunity for increasing the engagement of Native Hawaiian students with astronomy and for training of students for entry into the astronomy profession specifically or into the STEM workforce more generally.[80]

In 1968, the University of Hawaii received a Master Lease from the State of Hawaii to manage the Mauna Kea Science Reserve (MKSR), a ~13,000 acre region surrounding the summit of Maunakea.[81] This lease will expire in 2033. It is anticipated that a proposal to negotiate continued land authorization will be brought before the Board of Land and Natural Resources in 2021 (this may have occurred by the time this report is published). Recently, an internal restructuring of the management of Maunakea has been approved by the Board of Regents of the University of Hawaii. Established in August 2020, the Center for Maunakea Stewardship is drafting a new master plan for the Mauna Kea Science Reserve to, “outline a vision for Maunakea that balances cultural practice, recreation, the unique educational and research opportunities and scientific discovery offered on Maunakea, with minimal disturbance to the mauna.”[82]

In 2020, the National Science Foundation issued a statement that “potential construction of TMT on Maunakea is a sensitive issue and plans to engage in early and informal outreach efforts with stakeholders, including Native Hawaiians, to listen to and seek an understanding of their viewpoints.”[83] In addition, the Governor of the State of Hawaii has issued a 10-point plan to guide future activities on the mountain, including plans for restricting the duration of future leases and for limiting the number of telescopes through decommissioning of some existing facilities, as also determined by the State of Hawaii Board of Land and Natural Resources.[84,85,86]

[78] See https://www.eso.org/sci/libraries/edocs/ESO/ESOstats.pdf.
[79] See https://uhero.hawaii.edu/wp-content/uploads/2019/08/UHERO_Astronomy_Final.pdf
[80] Current programs include https://maunakeascholars.com/ for high school students and https://www.akamaihawaii.org/ for undergraduates and professionals.
[81] See http://www.malamamaunakea.org/management/comprehensive-management-plan.
[82] See https://hilo.hawaii.edu/maunakea/.
[83] See https://www.nsf.gov/news/news_summ.jsp?cntn_id=301034&org=AST.
[84] See https://governor.hawaii.gov/newsroom/news-release-governor-david-ige-announces-major-changes-in-the-stewardship-of-mauna-kea/.
[85] See http://www.malamamaunakea.org/management/comprehensive-management-plan/decommissioning.
[86] See https://dlnr.hawaii.gov/occl/files/2019/08/3568-TMT-Final-Decision-and-Order.pdf.

BOX 3.3 The Model of Community Based Science

Community Archaeology is the practice of archaeological research in which "at every step in a project at least partial control remains with the community"[87] with an emphasis on continually and meaningfully addressing the questions, "Who has access to [the] research? Who benefits? In what ways?"[88] The goal is to engage in genuine, two-way dialogue between researchers and the affected public from the outset, thereby to understand the types of investment—political, social, and/or material—that will best empower involved communities and lead to research activities that are "engaged, relevant, ethical, and, as a result, sustainable."[89] It is important to note that "involved communities" here are not necessarily limited to legally sovereign entities. Indeed, the sense of "at least partial control" is best understood in precisely these situations; it is intended to be an empowering notion. Whereas recognizing and respecting sovereignty where it exists (such as with many Native American tribes) is mandatory, the model of community-based science calls for researchers not to insist on exclusive control even in situations where it would be legally permitted to do so.

There are now a number of successful case studies of Community Archaeology in practice that can serve as exemplars for what we might regard as a Community Astronomy approach. For example, the excavation at Cancuén, Guatemala, nearly 20 years ago worked with the local community to develop a shared governance model that sought to maximize both the intellectual and financial contributions that a major research project can make to the communities around the site.[90] Together with representatives of the local communities, the project created a research and community development plan that recognized the local people "as custodians of their own heritage." Recognizing that Indigenous people often have an interest in learning more about their own heritage and traditional ways of knowing, the community representatives were included in the selection of research projects at the site, were empowered to choose revenue generating projects of interest to them (e.g., ecotourism), and remained integrally involved in planning and stewardship of site preservation and restoration.

Additional examples of community-based approaches have been documented in other National Academies reports. For example, a recent report discusses ethical considerations for forestry research, finding that "some spiritual traditions understand entire forests, or individual trees within forests, as being sacred, inspirited, or of moral significance, and therefore as requiring respect or imposing duties," and that "depending on how biotechnology is understood by these indigenous communities, its use could be interpreted as violating the right to manifest, practice, develop and teach their spiritual and religious traditions, customs and ceremonies; the right to maintain, protect, and have access in privacy to their religious and cultural sites."[91] A report on Earth science research finds that "incorporating concepts like ethnogeology (how geological features are interpreted by cultures) into lessons can increase the accessibility of the Earth sciences. Presenting Earth sciences in a way that is commensurate with, rather than in opposition to, native perspectives of Earth systems has had some success and is worthy of [funding agency] education resources."[92]

An especially apt example for astronomy is from a polar research report on the development of an Arctic Observing Network (AON): An "inclusive vision of the AON is desired by many arctic residents who view their environment in a holistic way.[93] This committee respects that desire and acknowledges the importance and value" of "involv[ing] arctic communities in true partnership from the outset and

[87] Yvonne Marshall (2002) What is community archaeology?, World Archaeology, 34:2, 211-219, DOI: 10.1080/0043824022000007062.
[88] See https://www.ucpress.edu/book/9780520273368/community-based-archaeology, p. 12.
[89] Ibid.
[90] See https://science.sciencemag.org/content/309/5739/1317.
[91] See https://www.nap.edu/catalog/25221/forest-health-and-biotechnology-possibilities-and-considerations.
[92] See https://www.nap.edu/catalog/13236/new-research-opportunities-in-the-earth-sciences.
[93] See https://www.nap.edu/catalog/11607/toward-an-integrated-arctic-observing-network.

recognizes that the inclusion of local and traditional knowledge and community-based monitoring will require a significant new investment and appreciation of local language, multiple literacies, and intellectual property rights." The report further argues for "collaboration with local communities and incorporation of local and traditional knowledge (LTK)" but that this "will take significant investment of time and resources and careful consideration of proper communication, data collection methods, and access and control of information." The report concludes with the following recommendation: "Arctic residents must be meaningfully involved in the design and development of all stages of the Arctic Observing Network. From the outset, the system design assessment should cultivate, incorporate, and build on the perspectives of human dimensions research and arctic residents. The Arctic Observing Network must learn what is needed to facilitate the involvement of local communities and create an observing network that is useful to them as well as to scientists and other users."

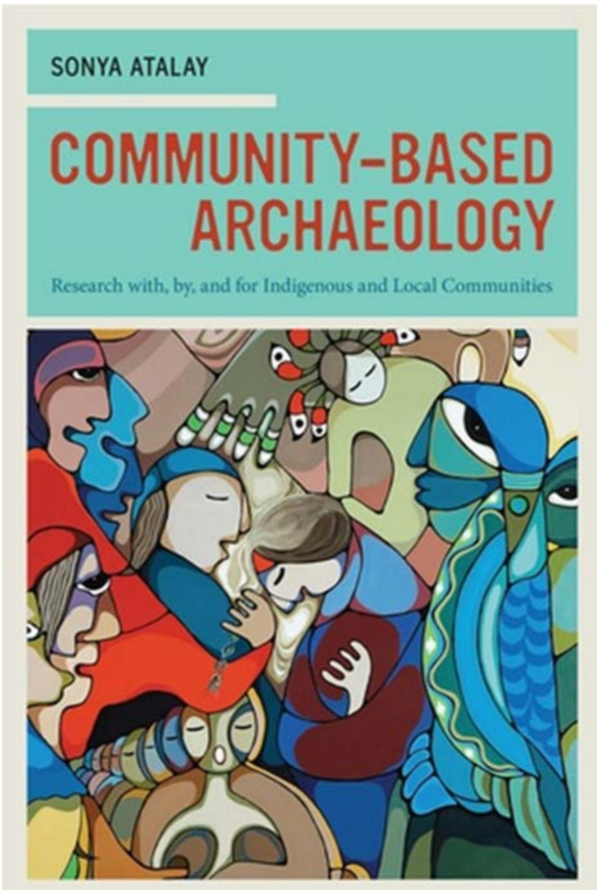


FIGURE 3.14 Community-based science—a model for research in which at every step in a project at least partial control remains with the community—is an approach that has been implemented to various degrees in archaeology, forestry, arctic science, and others. It can serve as an example for a Community Astronomy approach to active, up-front, and sustained engagement with local and Indigenous communities. SOURCE: Cover art by Daphne Odjig from S. Atalay, 2012, *Community-Based Archaeology: Research with, by, and for Indigenous and Local Communities*, University of California Press; reproduced with permission.

Astronomy can follow the example of archaeology, forestry, arctic science, and others to develop a Community Astronomy approach toward a more sustainable model of engagement with local and Indigenous communities. Such a community-based model requires first that the astronomy community

adopt and communicate a shared set of values and principles that guide it. These values for how to conduct ourselves and engage with one another include: Respect, Reciprocity, Trust, and Integrity.

To be sure, these are not the only shared values, and they are not unique to this matter. At a minimum these values speak specifically to both the failures of past engagements and to the healing required for positive sustained engagement going forward. To ensure alignment of current and future engagements with these shared values, the astronomy community could commit to the following principles specifically as part of a Community Astronomy model:

- *Listen and empower.* Make every effort to ensure all stakeholders are heard; while it may not be possible for all to have a formal say or vote in every matter, all can have a voice, and all stakeholder voices deserve to feel listened to. At the same time, a true community-based approach empowers the local community with at least partial control, even if power-sharing is not legally required (see Box 3.3); actively listening to the community means giving the community a seat at the table where decisions are made and where governance occurs.
- *Aim to do good for all.* The astronomy community adopts a higher standard than the bare minimum of legal compliance. Beyond the scientific benefits, astronomical activities would ideally add human value—educational, cultural, economic—respecting that different communities and cultures may ascribe value in different amounts or kinds, and may judge worth and worthiness through different lenses. A corollary is that the astronomy community must be willing to sometimes make difficult choices, and to be open to alternative solutions that optimize more than the science alone.
- *Invest in the future, together.* We cannot change the past, but we can make effort—extraordinary effort, if necessary—to work in partnership with communities and stakeholders to create a future defined by positive, long-lasting mutual benefit and respect for diverse ways of knowing. Communities and stakeholders are defined not by legal status alone but also by history, by potential impacts, and by opportunities. Regardless of the ground we stand on, we share a wonderment of one sky, and the quest for human understanding and connection with the cosmos can only be realized through full engagement of our diverse human talents.

Recommendation: The astronomy community should, through the American Astronomical Society in partnership with other major professional societies (e.g., American Physical Society, American Geophysical Union, International Astronomical Union), work with experts from other experienced disciplines (such as archaeology and social sciences) and representatives from local communities to define a Community Astronomy model of engagement that advances scientific research while respecting, empowering and benefiting local communities.

In support of this important goal, the astronomy community will need to seek to affirm, communicate, and continually reaffirm the astronomy community's framework of values and principles above for engagement with all stakeholders. The astronomy community could, as a sign of mutual respect, implement new journal citation standards, developed in partnership with Indigenous communities, that can be used in journal articles and talks in order to appropriately and respectfully credit Indigenous Traditional knowledge, oral histories, and protocols, and acknowledge the use of historically Indigenous lands. In addition, in alignment with other recommendations in this report toward increased transparency and accountability, facilities could engage in proactive efforts to assess local, societal, and cultural impacts—through a Community Astronomy approach that goes beyond mere regulatory compliance—including all stakeholders; as recommended in a previous National Academies report, "facility design should cultivate, incorporate, and build on the perspectives of human dimensions research."[94] Facilities could also report openly and regularly on these assessments, and make plans for

[94] See https://www.nap.edu/catalog/11607/toward-an-integrated-arctic-observing-network.

ongoing improvements, throughout the full life cycle of a project, that reflect the perspectives of all stakeholders. Finally, they would ensure that local stakeholders have meaningful influence—including through decision-making and governance structures at every stage—and involve local stakeholders in periodic assessments of when to decommission facilities.

In conclusion, there is the example of Arecibo Observatory, which at the time of this writing experienced an unexpected and catastrophic loss due to a support cable failing and leaving a 100-foot gash in the dish below and collapse of the platform and towers. In November, 2020, NSF announced the beginning of its plans to fully decommission the facility.[95] The observatory has, over the course of its nearly 60-year history, become very highly regarded by many of Puerto Rico's citizens, as a source of pride and local economic benefit, as well as of access to training and employment for many local people. Already, there is a groundswell of local support for efforts to preserve the site for educational and cultural activities even if not for research; recognizing the challenges of maintaining the visitor's center while the future is being planned, Astro2020 supports its continuation as an important nexus for education, community, and developing a diversified STEM workforce. The future of the Arecibo site for scientific research is discussed further in Section 5.1.5. As in the case of Maunakea and other sites, a Community Astronomy approach could fruitfully guide NSF, the local community, and the astronomy community in making plans for the disposition and future manifestation of Arecibo in a manner that is consistent with the scientific and programmatic priorities of this decadal report and that reflects the values and principles articulated above.

> **Conclusion:** NSF, NASA, DOE, facility managing organizations, project consortia, individual institutions, and other stakeholders can work to build partnerships with Indigenous and local communities that are more functional and sustained through a Community Astronomy approach, and by increasing the modes of engagement and funding for: (i) meaningful, mutually beneficial partnerships with Indigenous and local communities, (ii) culturally supported pathways for the inclusion of Indigenous members within the profession, and (iii) true sustainability, preservation, and restoration of sites.

3.4.2 Light Pollution and Radio Frequency Interference

The sensitivity of ground-based optical telescopes has been impacted by human-made light pollution for more than a century. The search for darkness has driven new observatories to remote sites, while pursuing local regulations to mitigate light pollution and interference. Nonetheless, increasing human population density and new technologies such as light-emitting diode (LED) fixtures continue to encroach on major observatory facilities (including the Vera Rubin Observatory). The collection of radio frequency data for astronomical use has had impacts from sources of radio frequency interference almost from the origins of radio astronomy. Recent developments in technology designed to improve quality of life such as car radar and radio frequency identification tags among others, increase the amount of radio frequency interference experienced by radio astronomers. Satellite constellations pose a parallel threat to the radio sky as to ground-based optical telescopes.

3.4.2.1 Light Pollution from Satellite Constellations

In the coming decade a new technological advancement threatens ground-based optical observatories. Earth-orbiting satellites have always been visible to astronomical telescopes (and human eyes), but their numbers were small enough they had minimal scientific impact. The situation is rapidly changing. Vastly reduced satellite launch costs, effective networking technologies, and ambitions for

[95] See Section 5.1.5 for a detailed discussion of the Arecibo Observatory.

global low-latency data-transmission have advanced plans for so-called "megaconstellations." Since mid-2019 when the Astro2020 Decadal Survey process began, the number of large (>100 kg) satellites in low Earth orbit has increased by an order of magnitude, and this extremely rapid growth is likely to accelerate. At the time this report was being prepared, three major constellations (SpaceX Starlink 1st and 2nd generation, OneWeb Phase 2, and Amazon Kuiper) were being proposed with a total of tens of thousands of satellites in low Earth orbit.[96] This landscape is evolving very rapidly, but the threats to nighttime astronomy as well as radio astronomy (Section 3.4.1.2 below) are clear.

Finding: Under current proposals the number of large low Earth orbit satellites will increase by orders of magnitude compared to 2018 levels, owing to reductions in launch costs, expected increasing demand for internet connectivity, and increasing effectiveness of networked satellites.

Spacecraft in low-Earth orbit also experience contamination from these satellites crossing their field of view,[97] and this is an important consideration for space assets in this region in the present and future. The topic of space debris is one that has long had the attention of NASA, and the increased number of satellites in these megaconstellations will almost certainly affect collision frequency. While these are still of concern, the new threat to the dark skies of ground-based optical astronomy is the one that requires assessment due to the new and changing environment.

Thousands of these satellites will be easily detected by modern telescopes, with their brightness depending on the time of night, position above the horizon, and on the phase of the satellite's life cycle. The satellites are only visible at visual wavelengths when sunlit, and thus are most visible during twilight, or at low elevations looking in the sunset or sunrise direction. Higher-altitude satellites may be somewhat fainter but will remain in sunlight longer and at higher elevations, and hence have larger impacts; during summer, some satellites at 1000 km altitude may be visible through the entire night.

A satellite's brightness varies both with distance and the satellite's orientation. The visually-striking Starlink "trains" occur only for the early phase of a satellite's life-cycle, after a group of satellites has been deployed and while they are being raised to operational altitude. During this phase, large surfaces such as solar panels are visible from Earth and the satellites may be comparable in brightness to naked eye stars seen in twilight. In the operational phase, the satellites are no longer concentrated in a single train, and their solar panels are oriented towards the Sun; reflections off the satellite lower surface are fainter (but often still visible to the naked eye), with a prospect of reducing that further. Other satellites will likely show similar behavior depending on their design details.

Wide-field imaging survey telescopes such as the Vera Rubin Observatory suffer the most severe impacts from megaconstellations. Its large field of view increases the probability of a satellite being present, particularly for science programs that are executed in twilight and at low elevation such as near-Earth object (NEO) searches for asteroids. Impacts may also be significant for programs that rely on extreme control of systematics such as large cosmological weak-lensing surveys. The large aperture of Rubin Observatory means that the satellite will approach or exceed the saturation level of the detector. Cross-talk between different detectors in the camera will result in multiple ghost trails whose brightness is a nonlinear function of the main satellite trail. Modeling by the Rubin Observatory project indicates that some of these effects might be mitigated by data processing, particularly if satellite brightnesses are reduced,[98] but overall may have a significant impact on many science programs.

Conclusion: The impact of megaconstellations is noticeable for wide-field imaging at optical wavelengths, will become more significant in the future, and will be potentially severe for some

[96] Venkatesan (2020) https://www.nature.com/articles/s41550-020-01238-3.

[97] https://indico.esa.int/event/370/contributions/5925/attachments/4238/6337/Sandor_Kruk_The_impact_of_satellite_trails_on_Hubble_observations_compressed.pdf.

[98] Tyson (2020) https://ui.adsabs.harvard.edu/abs/2020AJ....160..226T/abstract.

programs (especially from satellites in orbits above 600km), unless their effects are mitigated. Facilities especially impacted include the Vera Rubin Observatory. Scientifically the greatest impact is on searches for Near Earth Objects.

Assessing impacts and possible regulatory frameworks are now being addressed on a number of fronts by government agencies and the international astronomical community. This is a dynamic situation with complicated regulatory aspects; even in the last year, the situation has changed dramatically. JASON (an independent group of academic leaders that interfaces with the security community) was asked by NSF and DOE to assess the impact of current and planned large satellite constellations on astronomical observations, and issued reports in September and November 2020.[99] The AAS has formed a Working Group on Satellite Constellations, and co-sponsored workshops with the NSF NOIRLab (SATCON1) in the summers of 2020 and 2021; even more recent efforts (SATCON2) took place towards the end of the period of this decadal study. Since this is a global threat, the AAS has also coordinated closely with the International Astronomical Union (IAU) in addressing the issue. In May 2021, the IAU presented a Conference Working Paper to the Scientific and Technical Sub Committee of the United Nations Committee on the Peaceful Uses of Outer Space. The approach of these groups has largely been to facilitate dialogue between the astronomical community, the relevant aerospace companies, and national and international stakeholders. A public fact-finding workshop was held by this decadal survey on the issue, and it was attended by representatives of one such company, SpaceX.[100] SpaceX has been responsive and has been exploring methods for reducing the impacts of their constellations.

Addressing this growing challenge will require the same levels of coordination and ongoing attention by the astronomical community and agencies that has served the radio astronomy so well over the past decades. This need includes providing accurate models of satellite visibility and impacts, coordinating between astronomers and satellite operators, developing mitigation approaches, and advocating for astronomy. The entry of the NOIRLab to this arena is especially welcome, and the survey committee envisages it playing a similar coordinating role to the one that NRAO has fulfilled so effectively in radio spectrum protection. It is crucial that this framework be developed soon, so that mitigations can be built in during the early stages of constellation design and deployment. It is beyond the scope of this survey to recommend specific actors and actions, particularly due to the dynamic evolving nature, but it is clearly an issue that requires broad participation.

Recommendation: The National Science Foundation should work with the appropriate federal regulatory agencies to develop and implement a regulatory framework to control the impacts of satellite constellations on astronomy and on the human experience of the night sky. All stakeholders (U.S. astronomers, federal agencies, Congress, satellite manufacturers/operators, and citizens who care about the night sky) should be involved in this process. This is an international issue; therefore, international coordination is also vital.

3.4.2.2 Radio Frequency Interference

Threats to the radio sky differ from those in optical and infrared astronomy. Radio frequency interference (RFI) is multidirectional, and radio services, including commercial, military, and scientific operators, share the same spectrum. The system is managed by spectrum allocations to the various interests. There is increasing pressure on the radio spectrum from commercial interests, particularly at high frequencies that were previously of interest only to radio astronomers.

[99] Impacts of Large Satellite Constellations on Optical Astronomy, JSR-20-2H-L2, September 10, 2020. Space Domain Awareness: Impacts of Large Constellations of Satellites, JSR-20-2H, November 2020.
[100] April 27, 2020.

The radio spectrum, defined as electromagnetic radiation up to 3 THz, is coordinated internationally by the International Telecommunications Union (ITU), an agency within the U.N., which proposes intergovernmental treaties on the coordination of spectrum. Allocations for radio astronomy form a small portion of the available spectrum: only ~1.5 percent at frequencies less than 5 GHz (6 cm), 29 percent for frequencies less than 94 GHz (3 mm), and 65 percent in the range 95-275 GHz.[101] For the most part, radio astronomy is a passive user of the spectrum, with the exception of radar astronomy, which is primarily used for solar system observations. Modern, sensitive receivers seeking to detect faint sources use large bandwidths that are broader than the allocations specific to radio astronomy. Extensive observations of highly Doppler shifted radiation, such as galaxies in the early universe, means that frequencies are often shifted from their laboratory values, and lines can be at many locations in the spectrum. Frequencies of 90-240 GHz are used by projects such as CMB-s4, since they are near the peak of the cosmic microwave background spectrum. Modern, sensitive receivers working at these frequencies, particularly those employing broadband bolometric detectors, are vulnerable to RFI and cannot easily avoid or excise it.

Within the United States, the spectrum is managed jointly by the Office of Spectrum Management of the National Telecommunications and Information Administration (NTIA), within the Commerce Department, for federal interests, and by the Federal Communications Commission (FCC), for commercial interests. NSF is responsible for spectrum management for scientific purposes, through its Electromagnetic Spectrum Management Group (NSF ESM). This intra-agency group coordinates with the NTIA, the FCC on all aspects of spectrum management. The ESM also represents the U.S. internationally at meetings of the World Radiocommunication Conference (WRC). The National Academies Committee on Radio Frequencies (CORF) considers the needs for radio frequency requirements and interference protection for scientific and engineering research, coordinates the views of the U.S. scientists, and acts as a channel for representing the interests of U.S. scientists.

These regulatory and advisory structures have served the radio astronomy community relatively well. This section highlights two recent developments which will require close attention and management over the coming decade, namely the rapid expansion of the commercial broadband spectrum and RFI from satellite constellations.

Commercial services such as the mobile broadband standard 4G/LTE previously operated at frequencies below 1 GHz, and the resulting RFI issues were in the centimeter wavelengths. In spring, 2020, the FCC held an auction for allocations in 5 bands between 24 and 47 GHz, prime observing bands for the Very Large Array; these are likely frequencies for 5G technology. In addition, the FCC has recently stated that "The agency is creating new opportunities for the next generation of Wi-Fi in the 6 GHz and above 95 GHz band."[102] Mobile devices and smart vehicles will become widespread, moving sources of RFI. The higher radio frequencies, 20 GHz and above, are extremely valuable to astronomers. This frequency range figures prominently in the science case for the next generation Very Large Array, and is needed to accomplish science objectives like exploring the formation of solar system analogs on terrestrial scales, and using pulsars in the center of the Milky Way Galaxy for fundamental tests of gravity. At present, frequencies above 275 GHz are not controlled, and these are prime observing bands for the Atacama Large Millimeter/submillimeter Array (ALMA). Further encroachment into this band could impact ALMA science.

The new existential threats to radio astronomy observatories are satellite constellations. Instead of a limited number of satellites in relatively predictable orbits in the geostationary orbit, which can be avoided, the new trend is for constellations of low Earth orbit satellites. In addition to downlink radio signals, there are also inter-satellite radio signals for station-keeping. The proliferation of these satellites will render spatial avoidance of RFI extremely difficult. To give one example, the constellations of satellites from Space X's Starlink and OneWeb pose a significant risk to measurement of the CMB in the 20 and 40 GHz bands if steps are not taken to turn off transmission when the transmission beam and its

[101] van Zee presentation to the steering committee, 9 June, 2020.
[102] https://www.fcc.gov/5G.

sidelobes overlap observing sites. Planned CMB experiments have fields of view between 9 deg and 35 deg wide, and they achieve their sensitivity by measuring all the power that lands on them in roughly a 30 percent bandwidth. At any time, there will be multiple satellites in their field of view, and the satellite's RF power at peak transmission will blind the detectors. Even when they are in the sidelobes, their emission will be significant. An additional concern is the frequency purity of the signal. Second, third, and fourth harmonics are in other key observing bands from 85-105 and 140-170 GHz. Without action, RF emission from these satellites may well eliminate bolometric measurements of the CMB, both in temperature and polarization from the ground in these critical frequency windows in the not-too-distant future.

Conclusion: The impact of commercial services and satellite constellations on radio frequency interference is becoming severe, and threatens the scientific study of cosmic microwave background radiation, as well as detections of faint continuum sources necessitating wide bandwidths. Future large facilities especially impacted are the CMB-S4 and ngVLA; in particular, the lower frequency bands of the CMB-S4 project will be compromised and may become unusable unless action is taken.

To protect access to the radio sky, sources of RFI need to be eliminated to the greatest extent possible (see Figure 3.15). Mitigation through post-observation software analysis is not always possible, since very bright sources of RFI, unplanned out-of-band emissions, or RFI that is broadband or slowly varying in time are difficult or impossible to excise with software. Direct and early coordination between commercial, federal, and radio astronomy interests is critical, preferably with primary allocations for radio astronomy in key frequency bands. NSF is a key player in this process but DOE and NASA projects are also impacted. CORF has stressed the importance of spectrum management to radio astronomers and for the protection of radio observatories. "[D]eveloping coordination agreements between commercial applications (including satellites) and radio observatories is a critical step toward protecting radio astronomy receivers from direct transmissions that not only corrupt observations but could also damage equipment."[103]

In addition to ensuring allocations to critical bands for radio astronomy at frequencies of 95 GHz and above, passive use of the remaining spectrum by radio astronomers may be maximized through a multifaceted approach of careful spectrum monitoring and effective RFI mitigation. Strategies for mitigation include geographical separation, spectral separation, and temporal separation, and/or the establishment of a radio quiet zone. It is important that new facilities take account of the changing RFI environment, and the necessary RFI excision methods, when selecting a site and budgeting for hardware and software needs.

Finding: The radio frequency spectrum is a resource facing rapidly growing demands from commercial users such as satellite constellations and increased commercial use of higher frequencies, while at the same time new scientific instruments and capabilities increase the portions of the spectrum radio astronomers are using. Increasingly sensitive detectors can pick up on additional sources of interference.

Recommendation: To ensure that the skies remain open to radio astronomy, the National Science Foundation (NSF), in partnership with other agencies as appropriate, should support and fund a multi-faceted approach to the avoidance and mitigation of radio-frequency interference. It is critical that the astronomical community formally monitor commercial and federal uses of the spectrum managed by the Federal Communications Commission and the National Telecommunications and Information Administration and actively participate in the spectrum management process by seeking critical primary

[103] Astro2020 WP, Van Zee et al.

allocations to radio astronomy in the high-frequency bands above 95 GHz, by providing comments to filings for spectrum allocations, and by supporting the efforts of the Committee on Radio Frequencies, the National Radio Astronomy Observatory, and the Electromagnetic Spectrum Management division of NSF. To be most effective, international coordination is required.

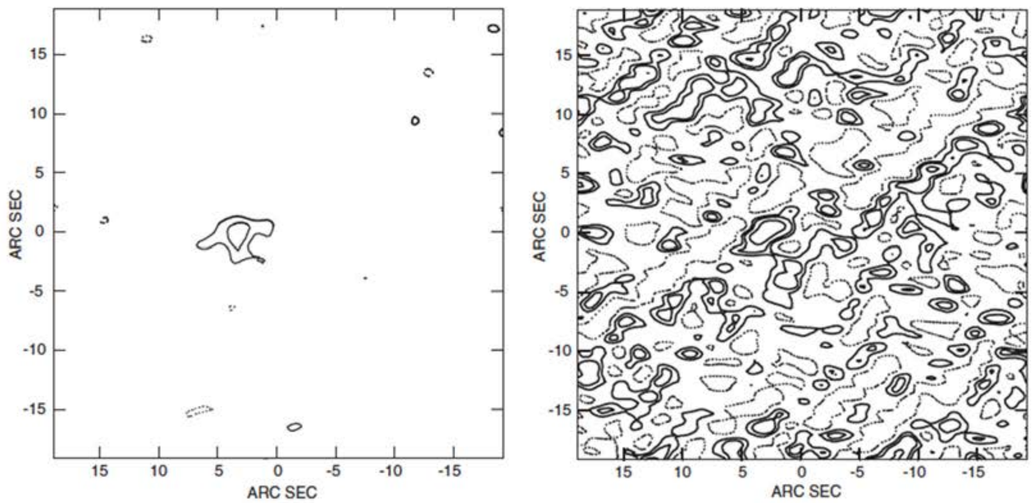


FIGURE 3.15 Effects of radio frequency interference on imaging at radio wavelengths with the VLA. Both images contain a faint radio star in a spectral window at 1612.22 MHz, a band in which radio astronomy has the primary frequency allocation; the satellite is producing spurious radiation in this band. There is no interference on the left, while the image at right was obtained when an Iridium satellite was 22 degrees from the star. The emission from the satellite swamps the extraterrestrial signal, rendering the data useless. SOURCE: G.B. Taylor, from National Research Council 2010. Spectrum Management for Science in the 21st Century. Washington, DC: The National Academies Press. https://doi.org/10.17226/12800. Courtesy of G.B. Taylor, NRAO/AUI/NSF.

3.4.3 Climate Change

As the twenty-first century progresses, human-induced climate change will be one of the greatest challenges. As with every other part of our society, astronomy and astrophysics must engage with this through several challenges: educating and informing people about this, understanding and minimizing our impacts on the climate, and recognizing and adapting to inevitable changes.

As noted in Section 3.2, individuals with training in astronomy and astrophysics are generally very strongly positioned for careers and leadership roles in science and technology beyond astronomy, and this includes specifically efforts toward climate change solutions. Indeed, astrophysics provides a natural home to discuss the greenhouse effect and global climate change. Greenhouse trapping of heat by increased mid-infrared opacity is a consequence of the same physics that determines the structure of stellar atmospheres. Our own solar system provides a natural laboratory to explore this concept, through the comparative temperatures of planets with varying abundances of CO_2 and other atmospheric heat traps; showing that Earth must be significantly warmed by greenhouse effects is an easy calculation for an introductory astronomy class. Studies of the solar cycle have helped confirm that external effects are not driving the warming observed over past decades. As exoplanetary systems are characterized, the greenhouse effect will be similarly important in their habitability.

Finding: Introductory astronomy classes could allow students to quantitatively understand the basics of global warming, and astrophysicists everywhere can be part of the public conversation reinforcing the reality of climate change.

As with other people and activities, astrophysicists also contribute to climate change. Two recent studies have shown that the professional activities of a typical astrophysicist generate ~20 - 35 tons of CO_2 per year excluding personal consumption such as food or home energy use.[104] This compares to approximately 20 tons per year for an average American including all sources. A significant contributor to the difference is air travel, along with emissions associated with electricity consumption from computation resources, particularly supercomputing facilities. Reducing this impact is an achievable goal for astronomy (as for all other fields).

Recommendation: The astronomy community should increase the use of remote observing, hybrid conferences, and remote conferences, to decrease travel impact on carbon emissions and climate change.

3.5 BUDGETARY IMPLICATIONS

The preceding sections of this chapter argue for a sustained recommitment to the future of the field, through significant re-investment in the profession and with an increased focus on matters of equity, diversity, and sustainability. For the astronomy and astrophysics profession, the benefits of these investments include: a workforce that, through its diversity, is more creative and innovative and reflective of society's full human potential; a professional community that, through equity and fairness, delivers on the promise of equal opportunity for all who would contribute their talent; and a set of policies and practices that, through their sustainability and accountability, ensure good stewardship of the natural and human resources necessary to achieve the field's ambitious science goals. For the broader society, the benefits of these investments include: expanded gateways to a very broad array of STEM careers; engagement in the excitement of astronomical discoveries for learners of all ages; expansion of the societal imperative of STEM literacy; and technological innovations with applications to remote sensing, navigation, and national security, among others. Together, these benefits contribute significantly to the nation's global leadership in science and technology beyond the obvious contributions to astronomical discovery.

The necessary investments span a range of types and costs. Indeed, a number of urgent recommendations can be implemented at little-to-no cost, such as policies and procedures aimed at combating racism, bias, harassment, and discrimination, or reducing the carbon footprint of professional activities. Some needs may already be addressed by current programs at the agencies. For example, NASA's PI Launchpad Workshop, held at U. Arizona in 2019, targeted diverse potential new NASA mission PIs. Still others will require non-trivial levels of funding, some new, some of it a restoration of previous investments. In Table 3.4 we provide budgetary guidance on those recommendations that carry funding implications for the agencies, drawing principally from the analysis and guidance provided by the SoPSI panel report. They are intended to provide rough guidance on the funding implications for meaningful action on our recommendations, and as a reminder to the community as a whole that such action requires investments. In keeping with the general approach of this survey we have refrained from dictating explicit programmatic priorities in general, in order to afford the agencies flexibility in obtaining and allocating the relevant funding. However the maintenance of accurate data on funding outcomes is sufficiently critical to the other recommendations that it is the most urgent need. The committee appreciates that stewardship of these important areas resides at various levels within the agencies, and may require coordination across them.

---

[104] Stevens et al 2020, Janke et al 2020.

TABLE 3.4 Budgetary Guidance Pertaining to the Profession and Its Societal Impacts

Recommendation	Funding Guidance[105] (annual)	Assumptions (see also SoPSI panel report)
Collecting, evaluating, and regularly reporting demographic data and indicators pertaining to equitable outcomes	\$0.5M - NSF \$0.5M - NASA	Modeled on effort at NIH.
Faculty diversity, early-career faculty awards	\$1M - NSF \$1M - NASA \$0.5M - DOE	Typical early-career faculty award of \$1M over 5 years.
Workforce development/diversity, “bridge” type programs and MSI partnerships	\$1.5M - NSF \$3M - NASA	Typical NSF PAARE site award of \$2.5M over 5 years; NASA MUCERPI site award of \$3M over 3 years.
Undergraduate and graduate “traineeship” funding	\$1M - NSF \$1M - NASA \$1M - DOE	Typical NIH T32 site award of \$1M over 3 years.
Independent postdoc fellowships	\$0.5M - NSF \$0.5M - NASA	Typical NASA Hubble and NSF AAPF awards of ~\$100k per year.
Mitigation of radiofrequency and optical interference from sources including satellite constellations	TBD - NSF	
Totals	\$4.5M - NSF \$6M - NASA \$1.5M - DOE	

3.6 CONCLUDING REMARKS

This chapter ends where it began, quoting from the SoPSI panel’s report: “The pursuit of science, and scientific excellence, is inseparable from the humans who animate it.” Indeed, the ability of astronomy and astrophysics to inspire and to awe is not only because of the grandeur of the Cosmos and the grandness of our wonderment about it; it is also, perhaps even more so, because it is people—seemingly so small and insignificant in relation to that vastness—who dream the questions and who dare to try to answer them. Our ability to grasp the universe is as great as it is because it is driven by the boundlessness and breadth of human curiosity, creativity, ingenuity, and diversity. The profession of astronomy and astrophysics understandably takes considerable pride in its many contributions to the nation and the world, not only to scientific knowledge but also as a shining example of how science can enrich, inspire and stir the imaginations of people everywhere, of all ages and walks of life.

At the same time, because it is a human endeavor, astronomy and astrophysics is not immune from human foibles and failings, nor wholly separable from larger societal forces—for better and for

[105] Amounts listed represent new funding, reinstated funding, or augmentations over current funding, as appropriate.

worse. This Astro2020 decadal exercise has been no exception. As aptly noted by the SoPSI panel (paraphrasing): As this report was written in mid-2020, the United States was in the midst of profound self-examination of social and economic inequalities resulting from historic and systemic racism, highlighted by the Black Lives Matter movement, sexual harassment and inequalities highlighted by the #MeToo movement, and the starkly inequitable and severe health and economic impacts of the COVID-19 pandemic on people of color, including a shocking and disturbing rise in violent crimes of hate against Asian Americans. For these reasons, the time during which this report was written was a dark time indeed for many in the astronomy community and around the world. Unfairly, it was an even darker time for those to whom fairness has too often been a thing denied.

Over the past decade our profession has made strides, individually and collectively, to address its longstanding structural inequities, borne of the historic barriers of race, gender, class, background, and identity inherited over decades across all of academia and society. As documented in this chapter and in the SoPSI report, slow progress is being made on many fronts, and through the leadership of the American Astronomical Society, the American Physical Society, and the American Institute of Physics, efforts are ongoing to build on the successes. Many of the major federally-funded institutes and NASA and NSF themselves are recognizing the needs and opportunities for leveling the playing field and removing the vestiges of bias and barriers to access in the awarding of resources. And the makeup of the field has become measurably more diverse, at least in some ways. These important steps are a beginning, and they are to be celebrated.

Against that backdrop it can be unsettling to many to be reminded that astronomy and astrophysics, like nearly all of the other sciences, still has a very long way to go before we can claim any semblance of victory over the inequities remaining within the system we oversee, regardless of how they came about, and the inordinate pressures that we often impose upon ourselves, especially among students, early-career scientists, and individuals from the many marginalized communities we represent and must encourage—including those discussed in detail above, as well as the disabled community, LGBTQ community, Muslim American community, and others—through the structure of our career pipeline and the environments we create in departments and workplaces. If we truly aspire to serve as a beacon and gateway to science for all people then our composition ought to reflect our people, all of them. If we aspire to create and nurture a professional family of individuals, we need to treat each other as family, with mutual respect, empathy, and support regardless of career stage, personal identity, or scientific identity within our diverse profession, and with no tolerance for those who abuse their position and their colleagues. And if we hope to continue to benefit from the resources of our planetary home and of the global communities that inhabit it, we must conduct ourselves with sustainability as a greater priority than ever before.

Much of the challenging task of exploring this complex landscape was taken up by the Panel on State of the Profession and Societal Impacts, and the report they produced was candid, and critical where it needed to be. Some will find passages to be provocative reading, whether the topic is racial and ethnic representation and discrimination, sexual harassment and discrimination, stewardship of observatory sites, or the many other issues and areas addressed in the report. Facing such truths by listening, reflecting, and facilitating ongoing dialog will uplift and empower not only those who face barriers to entering and advancing in the profession but also to enhance the entire astronomy and astrophysics community. It also requires the will to act, and a commitment to devote the resources necessary to ensure that our values are reflected not only in where we direct our labor but in how we spend the dollars entrusted to us. Together, the SoPSI report and this one strive for a common goal: to address our charge and provide constructive findings and conclusions—and to make actionable, resourceable recommendations—for making our profession more representative of our society, more inclusive, and a more collaborative partner with the communities within which we work. Which is to say, to make our profession better.

4

Optimizing the Science: Foundations

Building on Chapter 3, which describes the human investments and public impacts of the decadal survey's program, and how to enhance and diversify them, this chapter focuses on the essential scientific foundations, specifically the resource infrastructure and the underpinnings of how scientists turn ideas and data into discovery. This chapter draws from the Enabling Foundations panel report and identifies challenges and opportunities for progress over the next decade.

Astronomy and astrophysics have achieved breathtaking accomplishments over recent decades. Much of the credit for this sustained record of success can be attributed to diversified portfolios of investment by NASA, the National Science Foundation (NSF), and the Department of Energy (DOE), ranging from cutting-edge flagship observatories in space and on the ground to investments in mid-scale and smaller supporting facilities, and a foundation of support for the calibration, analysis, interpretation, and theoretical modeling of the rich data sets produced by these facilities. Although the decadal process primarily focuses on recommendations to the federal agencies that sponsor it, astronomy in the United States has also historically benefited from major investment by state universities, private universities, philanthropic foundations, and individual donors, in addition to the federal government. Private foundations and philanthropy play many roles in astronomy, from supporting observational facilities and large projects, to individual principal investigator (PI) support and postdoctoral fellowships (e.g. the Heising-Simons 51 Peg postdoctoral fellowships), and key seed funding for future projects. Over the last two decades private foundations are increasingly playing a key role supporting individual researchers, either directly through grants programs or through institutional support (e.g. the Kavli centers, Carnegie Observatories, or the Simons Center for Computational Astrophysics). For example, the Brinson, Guggenheim, Heising-Simons, Kavli, Moore, Packard, Research Corp., Simons, Sloan, and Templeton Foundations all provide significant funding to individual astrophysics researchers, including early career postdocs. This is critical support for U.S. astrophysics research in theory, computation, and support for archives in a time of increasing competition for federal funding.

This foundation of smaller-scale investments often receives less attention in decadal planning exercises, but its importance cannot be overstated. The most complex and precise measurements would mean nothing without pipelines to calibrate and process the raw data, algorithms to analyze and interpret the data, theoretical calculations to provide a context for the results and help understand their implications, and support for people who do the science. Laboratory astrophysics measurements, data science and computational methods, and data archiving all play critical roles in turning photons, particles, or waves into scientific insights. These foundational programs have the potential to bring more people into the field through reducing barriers to participation by anyone through supporting their success, and through offering access to state-of-the-art tools, training, and facilities. Such programs provide the seed corn for the future innovators and leaders in the profession.

4.1 THE IMPORTANCE OF A BALANCED PROGRAM

Federal astronomical research investments in the United States today can be subdivided into a few critical components. The largest federal investments are in major flagship observatories and facilities,

such as the original NASA Great Observatories (the Hubble Space Telescope, Chandra X-ray Observatory, Spitzer Space Telescope, Compton Gamma-Ray Observatory), and now the James Webb Space Telescope (JWST) and the Nancy Grace Roman Space Telescope, or, in ground-based facilities funded through the NSF Major Research Equipment and Facilities Construction (MREFC) line (e.g., Gemini, the Atacama Large Millimeter/submillimeter Array [ALMA], the Daniel K. Inouye Solar Telescope [DKIST], Vera Rubin Observatory, Laser Interferometer Gravitational-Wave Observatory [LIGO], and IceCube). A second critical component of the nation's observational capabilities is the suite of smaller-scale facilities and dedicated survey instruments, with examples of the former including NASA Explorer-class missions and examples of the latter including the Sloan Digital Sky Survey (SDSS). Many of these mid-scale facilities are funded and operated by partnerships between public agencies, private institutions, and foundations. Both classes of facilities have delivered breakthrough discoveries including Nobel Prizes, along with thousands of smaller individual and team-led investigations that collectively have fueled the extraordinary success and growth of the field in recent decades.

The other essential components of the national investment portfolio are the people who carry out and drive the science (addressed separately in Chapter 3), and the enabling foundation or infrastructure that supports research. The data produced by the aforementioned facilities would be lost without investment in this foundation: support for processing and archiving these data, analyzing and interpreting the data, theoretical modeling and simulations, and in many cases carrying out the laboratory and computational analysis of the atomic, molecular, and chemical signatures and diagnostics of the emitting processes. The aggregate investment in this foundational support comprises a small fraction of the overall agency portfolios, but it multiplies by several-fold the scientific yield from the facilities.

The need for balance across this broad portfolio has long been recognized by the agencies themselves. A prime example is the wide range of mission size classes supported by NASA through its Astrophysics Science Mission Directorate, ranging from flagships to small and medium Explorers, and extending down to small satellites, CubeSats, and support for balloon and rocket payloads. This healthy mix enhances scientific balance, and cost effectiveness. It also provides pathways for new and early-career investigators to build scientific, technical, and leadership experience with progressively larger mission classes. Another commendable example is the provision in NASA mission lines for the costs of data processing, analysis, and archiving, and in the case of flagship missions, support for guest observer grants, which encourage the timely analysis, publication, and dissemination of the data and science resulting from the missions, including theoretical studies needed to interpret the data.

The funding portfolio of the NSF Division of Astronomical Sciences (AST) is unlike that of most other NSF divisions, including those in the NSF Mathematical and Physical Sciences Directorate (MPS). Astronomy has a long history of capitalizing on shared, communal infrastructure that far exceeds the capability of any one individual institution, and that, for NSF facilities, is accessible by anyone. Under this model, a considerable fraction of the AST budget supports national observatories and facilities including the National Optical-Infrared Astronomy Research Laboratory (NOIRLab) and the National Radio Astronomy Observatory (NRAO), including U.S. participation in ALMA. These facilities support comprises about 75 percent of the total AST budget, with the remainder available to support individual investigator research, mid-scale infrastructure programs, and division-specific education activities. The fraction of research funding (<25 percent) is far lower than those of other MPS divisions (45-95 percent), but reflects in part the disproportionate number of shared national facilities in ground-based astronomy.[1] The current model through which MREFC projects are funded for operation by divisions has resulted in an unbalanced program in NSF's AST division that is not sustainable. The structural problem is addressed in Chapter 5 while this chapter focuses on balance in specific programs.

Although the primary role for advising agencies on their funding portfolios on an ongoing basis rests with the various agency and National Academies-administered standing committees, it is appropriate that the decadal surveys assess in broad terms the impacts of the current balances, and where appropriate,

[1] National Science Board, 2018, *Study of Operations and Maintenance Costs for NSF Facilities*, NSB-2018-17, Alexandria, VA, https://nsf.gov/pubs/2018/nsb201817/nsb201817.pdf.

to identify areas where a more healthy balance could be achieved. Here “healthy” is regarded as a balance of program investments that optimize the overall scientific productivity and future sustainability of the enterprise. This follows the practice of recent decadal surveys, and some of the findings and recommendations here will echo those already noted in the 2000 and 2010 surveys.

When addressing the question of balance, the Panel on an Enabling Foundation for Research and the steering committee identified a few critical areas where evolution in the funding balances within NSF AST and NASA Astrophysics has drifted into unhealthy territory, or, where the evolution of the research landscape itself has led to the need for enhanced investment in emerging disciplines. Chief among these are support for investigator grants for research and data analysis, and in the infrastructure support for data archiving, processing, and analysis as well as the related needs in computation, software, and data science. Previously identified under-investments in laboratory astrophysics and theory remain as critical needs. Each of these areas is addressed in the remainder of this chapter.

4.2 ENABLING SCIENCE THROUGH A HEALTHY INDIVIDUAL INVESTIGATOR GRANTS PROGRAM

It is people who are the source of American scientific and technical prowess, and supporting those scientists is the way to realize the scientific visions that are put forward in Chapter 2, A New Cosmic Perspective. Access to world-leading facilities is not enough to produce science. Individual scientists need access to the financial resources that allow them to collect, analyze, and interpret data from those facilities. That funding unlocks the effort of scientists and trainees to explore new ideas or to execute the hard but important projects that drive the field forward. Without resources, however, scientists’ insights and talent lie unrealized and discoveries unmade.

Chapter 3 emphasized the need to collect demographic information from researchers in external grant programs to assess indicators pertaining to outcomes of proposal competitions. A lack of data is apparent here as well; proposal success rates for only a few programs were available, and not always the most recent data. NSF noted that it is against their policy to release any information about proposals that have not resulted in awards; moreover, a recent policy prohibits the public release of proposal selection rates, so the number of submitted proposals and total request amounts were not made available to the survey. While NASA collects some data on proposers, the agency has only started to assess and evaluate it in a systematic way. Having these data would have better informed this report.

Conclusion: The lack of publicly available data on proposal success rates by program and on aggregated metrics for who and what type of research is being supported (theory, facilities, laboratory investigations, investigator demographics, student vs. postdoc funding for example) hampers analysis, evaluation, and oversight.

Recommendation: The National Science Foundation, NASA, and the Department of Energy should release data on proposal success rates on an annual basis, and should track metrics that allow them to analyze statistically what is being supported.

4.2.1 Bolstering the Individual Investigator Grants Program

Funding for the majority of astronomy research flows through “individual investigator grants,” where the lead scientist proposes a specific project and asks for the needed resources (salary for trainees, summer salary for senior personnel in academic positions, computing, travel, etc.) to bring the project to fruition. These proposals take a variety of forms. Ground-based astronomy research and theory is funded through the NSF Division of Astronomical Sciences, with research for “individual investigators” provided through its Astronomy and Astrophysics Research Grants (AAG) program. NASA funds research that is

relevant to space-based astrophysics missions, primarily through the Astrophysics Research and Analysis Program (APRA), Astrophysics Theory Program (ATP), Astrophysics Data Analysis Program (ADAP), and the Exoplanet Research Program (XRP). NASA also provides support for data analysis for U.S. investigators who have successful Guest Observer (GO) proposals on some of its active missions, as well as funding preparatory work for some future missions. There are also funding opportunities for theory and archival work connected to specific missions (currently Hubble, Fermi, Swift, and Chandra). In this NASA GO funding model, a successful research proposal for observing time is considered sufficient to unlock funding when the observing proposal is approved and executed. For ground-based NSF funded research, however, the funding is unconnected to awards of observation time, even on NSF-funded facilities.

The NSF AAG program is a cornerstone of the enabling foundation for research in astronomy and astrophysics in the United States. It supports research projects across nearly all subfields of the astrophysical sciences, and most of its funding supports individual investigators and their groups. Proposals are rigorously reviewed and the short funding durations for grants (typically 3 years) ensures that funding priorities reflect the most important scientific priorities in the field. The grants have led to discoveries that have transformed astronomy. For example, the work tracking stars in the immediate environment of the Milky Way Galaxy's black hole (Figure 4.1) began receiving NSF funding in the early part of the 2000s, and was recognized with a Nobel Prize in Physics in 2020.

Permission Pending

FIGURE 4.1 The change in position of two stars at the galactic center around what is now confirmed to be a supermassive black hole at the center of the Milky Way galaxy. This data was taken with the Keck telescope over a time period of 1995-2014, combining two decades of speckle imaging and adaptive optics data. This work was supported by NSF AST individual investigator grants. These improved mass and distance estimates were crucial for cementing the black hole explanation for Sgr A*, the subject of the 2020 Nobel Prize in Physics. SOURCE: Boehle et al. (2016), http://www.astroexplorer.org/details/apjaa2b70f5.

Preparing proposals for individual investigator grants is extremely time consuming, and—given the large impact a successful proposal has a large impact on a scientist's output and career—the stakes are typically high. NSF AAG proposal success rates averaged 30-50 percent in the early 1990's through the early 2000's (Figure 4.2), during which time most scientists had an expectation that their work could be funded within a reasonable time frame, perhaps after one or two resubmissions. However, funding rates

began to decline to below 30 percent beginning in the mid-2000's, and then even further to around 15 percent in the early part of the past decade. While funding rates have recovered somewhat, from 2010-2018 the average AAG success rate has remained around 18 percent, far below the 30 percent success rate target recently identified by NSF as a goal for the foundation overall.[2]

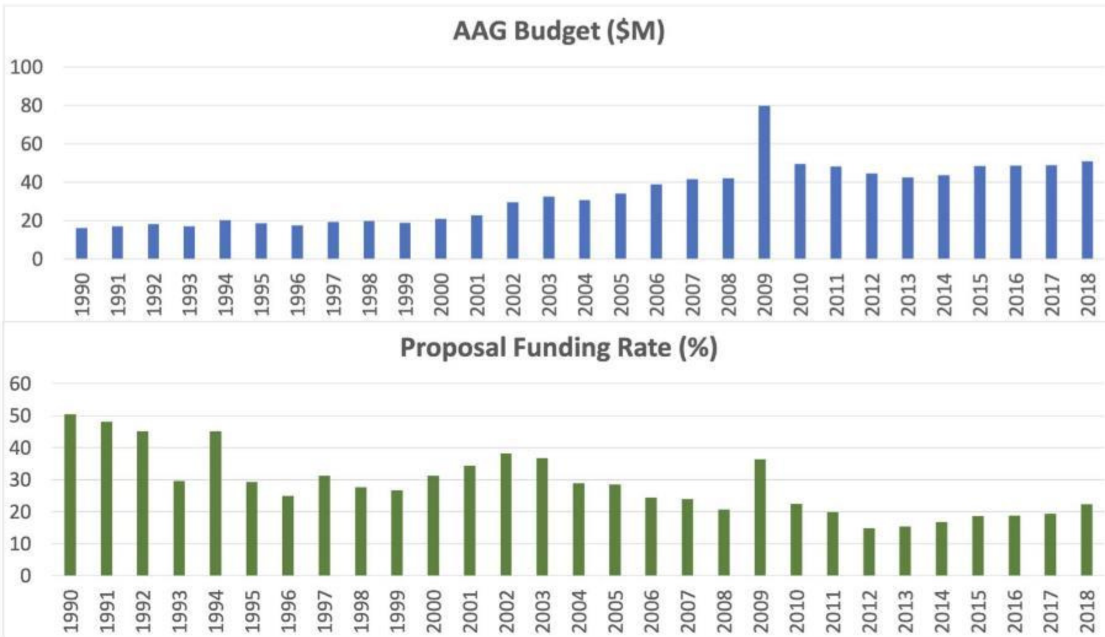


FIGURE 4.2 Plots of NSF AST Astronomy and Astrophysics Grant Budget and proposal funding rate versus time, from 1990-2018 in real-year dollars. The increase in funding in 2009 originated from the American Recovery and Reinvestment Act. SOURCE: Based on data from R. Gaume, National Science Foundation, presentation to the steering committee on July 15, 2019.

In 2015, a study group of the Astronomy and Astrophysics Advisory Committee (AAAC) investigated the impact of declining success rates at both NASA and NSF astrophysics programs on scientific productivity (Figure 4.3).[3] They concluded that the decline in success rates was not related to changes in the average proposal quality or to the fraction of proposals judged to be highly deserving of funding. Reviewers are instructed to grade proposals on an absolute scale from E (excellent), V (very good), G (good), F (fair), P (poor), where an Excellent proposal is an "Outstanding proposal in all respects; deserves highest priority for support" and a Very Good proposal is a "high quality proposal in nearly all respects; should be supported if at all possible." The fraction of proposals judged to be highly deserving of funding (VG, VG/E, E) has remained stable from year to year. However, the success rate of proposals ranked VG dropped from 45 percent in 2007-2008 to 25 percent in 2012. In other words, three out of four proposals that were judged as nearly certain to result in high quality science are rejected each year.

The AAAC group's quantitative analysis revealed that the major factor driving the increase in proposal oversubscription was that the budgets for these programs have not kept up with the increase in the number of (unique) proposers. This increase in the number of investigators tracks the overall addition

[2] https://www.aip.org/fyi/2021/panchanathan-makes-case-nsf-expansion-appropriators

[3] Cushman et al., 2015, "Impact of Declining Proposal Success Rates on Scientific Productivity," AAAC Proposal Pressures Study Group, https://arxiv.org/abs/1510.01647.

of researchers to the field, with no significant change to the mix of career stages over time. Individual budget items (typically dominated by salaries and tuition) have also increased in cost due to inflation, increasing the cost of funding a constant level of proposed effort.

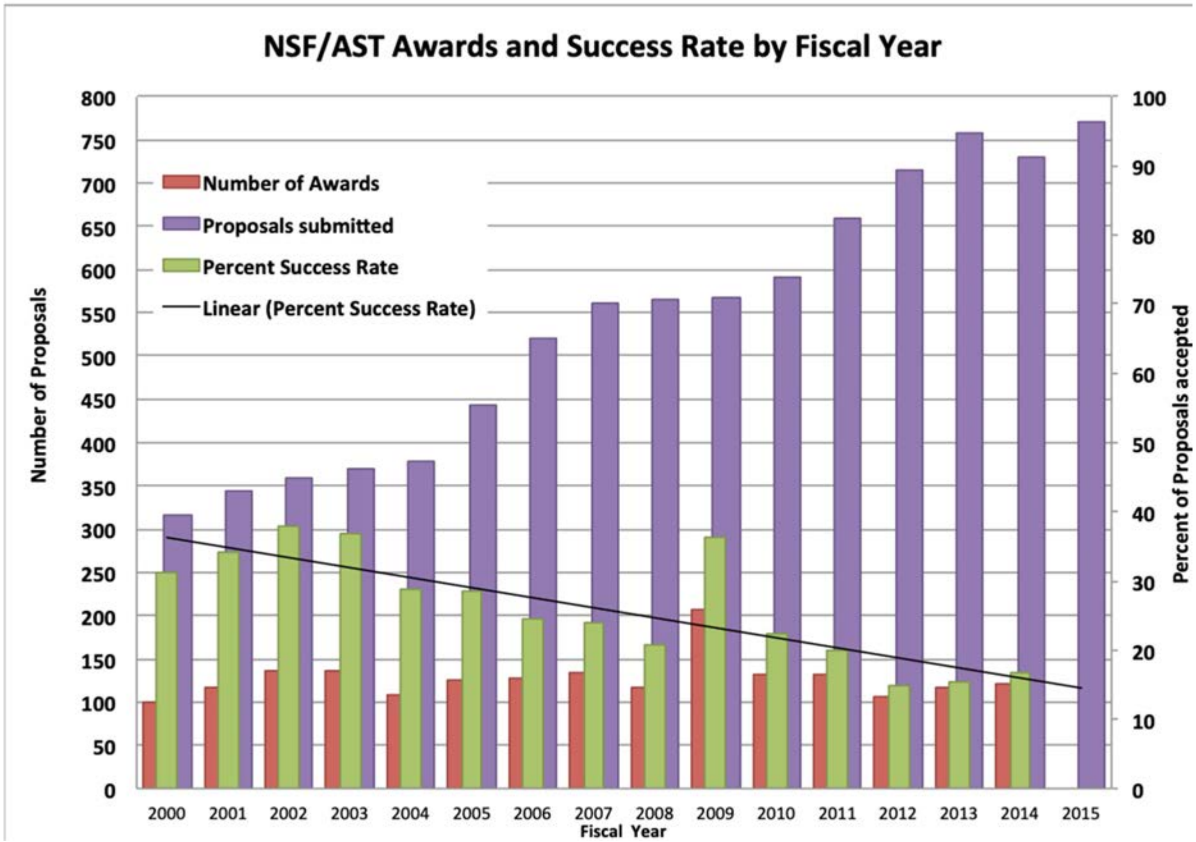


FIGURE 4.3 Historical NSF AST individual investigator grant statistics, from 2000-2015. There is a rise in the number of proposals submitted over this decade and a half, while the number of awards has not increased concomitantly. SOURCE: Cushman et al. (2015), Figure 1, https://arxiv.org/abs/1510.01647. Reproduced with permission.

The decrease in AAG proposal success rate is attributed by a 2018 National Science Board (NSB) report in roughly equal measure to the increase in the number of submitted proposals and the decrease in available funds because of the increase in facilities operations costs with a nearly flat AST budget.[4] AST stands out in the MPS directorate for having both a low proposal success rate and spending the least amount of its budget on individual grants programs.

> **Finding:** There is a systematic tension between funding facilities' operations and maintenance, and supporting the work of scientists able to turn data into discovery. The imbalance in AST has worsened over the last decade and will impact the ability to adequately support new facilities and new science going forward.
>
> This issue is developed in more detail in Chapter 5.

---

[4] National Science Board, 2018, *Study of Operations and Maintenance Costs for NSF Facilities*, NSB-2018-17, Alexandria, VA, https://www.nsf.gov/pubs/2018/nsb201817/nsb201817.pdf.

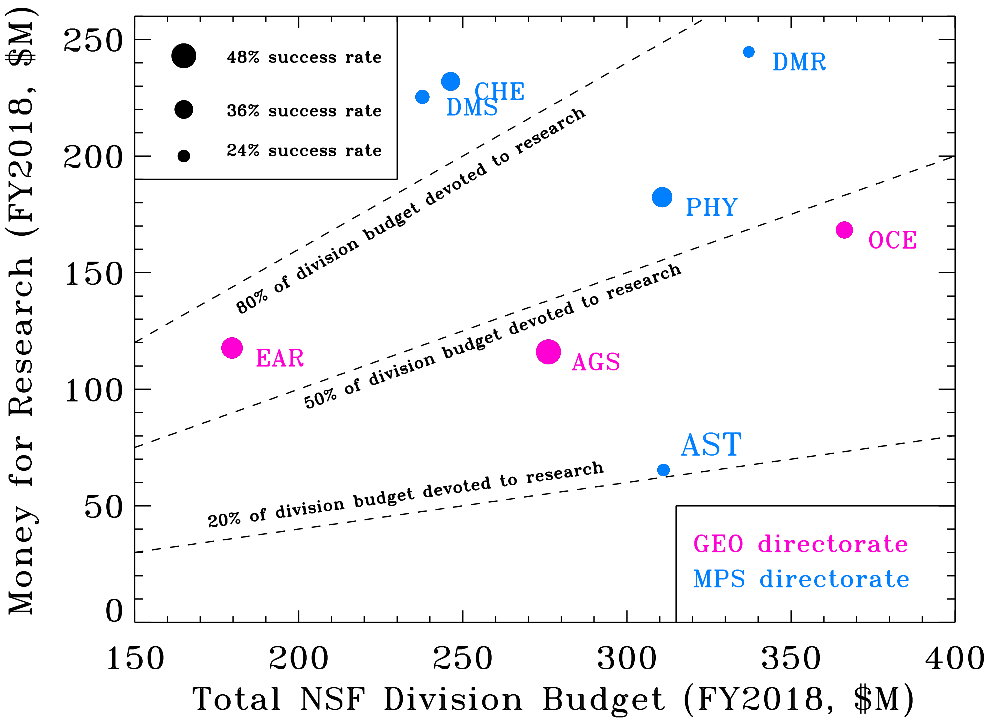


FIGURE 4.4 Money spent on grants to investigators as a function of the total NSF Division budget, for divisions within the Geosciences (GEO) and Math and Physical Sciences (MPS) directorates, for FY2018. The color coding indicates into which directorate the division falls, while the size of the symbol is proportional to the proposal success rate (in FY2020). Dashed lines indicate the fraction of the division budget devoted to research. The AST division stands out both for the low fraction of the division's budget used for research, as well as the low proposal success rate (24%). NOTE: AST=Astronomical Sciences; PHY=Physics; CHE=Chemistry; DMR=Materials Research; DMS=Division of Mathematical Sciences; OCE=Ocean Sciences, AGS=Atmospheric and Geospace Sciences; EAR=Division of Earth Sciences. SOURCE: R. Osten, NSF. based on data from https://www.nsf.gov/about/budget/fy2020/pdf/27_fy2020.pdf, https://www.nsf.gov/about/budget/fy2020/pdf/26_fy2020.pdf, https://www.nsf.gov/funding/funding-rates.jsp?org=MPS, https://www.nsf.gov/funding/funding-rates.jsp?org=GEO.

The AAAC report noted that the historical proposal success rates of 30-35 percent achieved for NSF AAG funding prior to and including FY 2003 was a healthy competitive environment, where the average proposer faced a manageable level of risk (~30 percent) of no funding after three attempts. Over the entire foundation, roughly 30 percent of proposals are ranked highly meritorious, and recent initiatives by the NSF Director are focused on achieving a grant proposal success rate of 30 percent.[5]

Other NSF divisions that share common features with astronomy, such as physics and oceanography both being heavy users of the MREFC line, have higher proposal success rates than the astronomy division and devote a larger fraction of their budget to supporting individual investigator grants (Figure 4.4). Differences in culture between different scientific fields may also contribute to this disparity, as some fields deprecate multiple proposal submission by a particular research group responding to a proposal call. Success rate alone is not the only factor to consider, as grants need to be of sufficient size to carry out the proposed science project.

[5] https://www.aip.org/fyi/2021/panchanathan-makes-case-nsf-expansion-appropriators

As pointed out in the Enabling Foundations for Research report, many programs within the NSF Physics Division (PHY) have success rates higher than 30 percent. Such levels also offer opportunities for early-career researchers proposing for the first time to have a realistic chance of success. The same AAAC study showed that when success rates dropped to lower levels first-time proposers fared even less well, with success rates as low as 7 percent. Chronic underfunding carries the additional risk of stagnation of the field.

The analysis above has focused on NSF grants, which are the sole way of publicly funding peer-reviewed science with ground-based telescopes. As noted in the Enabling Foundation report, while oversubscription rates for some NASA programs are healthier (e.g., APRA), support for other programs (e.g., ATP for theory) also appears to suffer from similar high proposal pressure and underfunding. The Enabling Foundation report suggested a 20 percent increase in funding above inflation for all individual investigator grants programs to restore success rates to a healthy competitive environment.

This underfunding has also impacted equity within the field. Ensuring adequate funding enables the whole community to reap the benefits from federally funded facilities. In the absence of public funding, scientists at wealthy institutions may still be able to tap into other sources of funds to support their research (while benefiting from institutional support for graduate students or endowed postdocs), or to carry out preparatory work to ensure successful proposals. However, these paths are largely out of reach for those at less affluent institutions, shutting them out of using the facilities the nation has invested in. This wastes decades of investment in, training of, and effort by flourishing scientists—by not providing the conditions in which they can execute their plans for science.

Conclusion: Robust individual investigator grant funding is crucial to meet the science challenges described in the Cosmic Ecosystems; New Messengers, New Physics; and Worlds and Suns in Context science themes. The historical proposal success rate for individual investigator grants within the NSF AST division of around 30 percent, realized at the start of the millennium for astrophysics programs, strikes an appropriate balance between a healthy competitive environment and a good chance of eventual success with resubmission. By any of several metrics which can be used to judge a healthy level of competition for individual research grants—dollar amounts spent on research, percentage success rates, fraction of high-quality proposals being rejected—the current state of individual investigator grants within the NSF AST division is not at a healthy level. Increasing grant funding is also required to ensure more equitable access to resources.

Recommendation: The National Science Foundation should increase funding for the individual investigator Astronomy and Astrophysics Research Grants by 30 percent in real dollars (i.e., above the rate of inflation) over 5 years from 2023-2028 starting with the fiscal year 2019 budget inflated appropriately. This will have the effect of restoring success rates to a healthy competitive level.

This funding augmentation is needed to reach the 30 percent proposal success rate goal, justified both from analysis of other programs and areas at NSF as well as being consistent with NSF's stated goal.

4.2.2 The Importance of a Healthy Theory Foundation

Theory is crucial in astrophysics, as both a mechanism for driving new discoveries and a framework for interpreting essentially all signals received from space. The focus of modern theoretical research has increasingly expanded from traditional pencil-and-paper calculations to complex computer simulations and sophisticated statistical analyses. Understanding prize-winning discoveries such as the Cosmic Microwave Background anisotropy and gravitational waves from merging black holes would not have been possible without the conceptual framework provided by theory. Indeed, breakthroughs in

theoretical predictions of the characteristics of the gravitational wave signal produced by black hole mergers were critical to the Nobel Prize-winning discovery based on LIGO detections (Figure 4.5). The science themes presented in Chapter 2 demonstrate that theory and observation are intertwined, necessitating a multi-pronged approach to addressing these important topics: a theoretical understanding of how gas of different temperatures and densities can co-exist in galactic outflows of differing velocities, for example, is essential to examining the processes that link matter inside of galaxies with its outside environs. As another example that relates to the priority science area of Pathways to Habitable Planets, theoretical calculations of planetary atmosphere chemistry and evolution will be needed to interpret biosignature gases detected in exoplanet spectra. This theoretical research lays the groundwork for designing new observational programs and planning for new facilities.

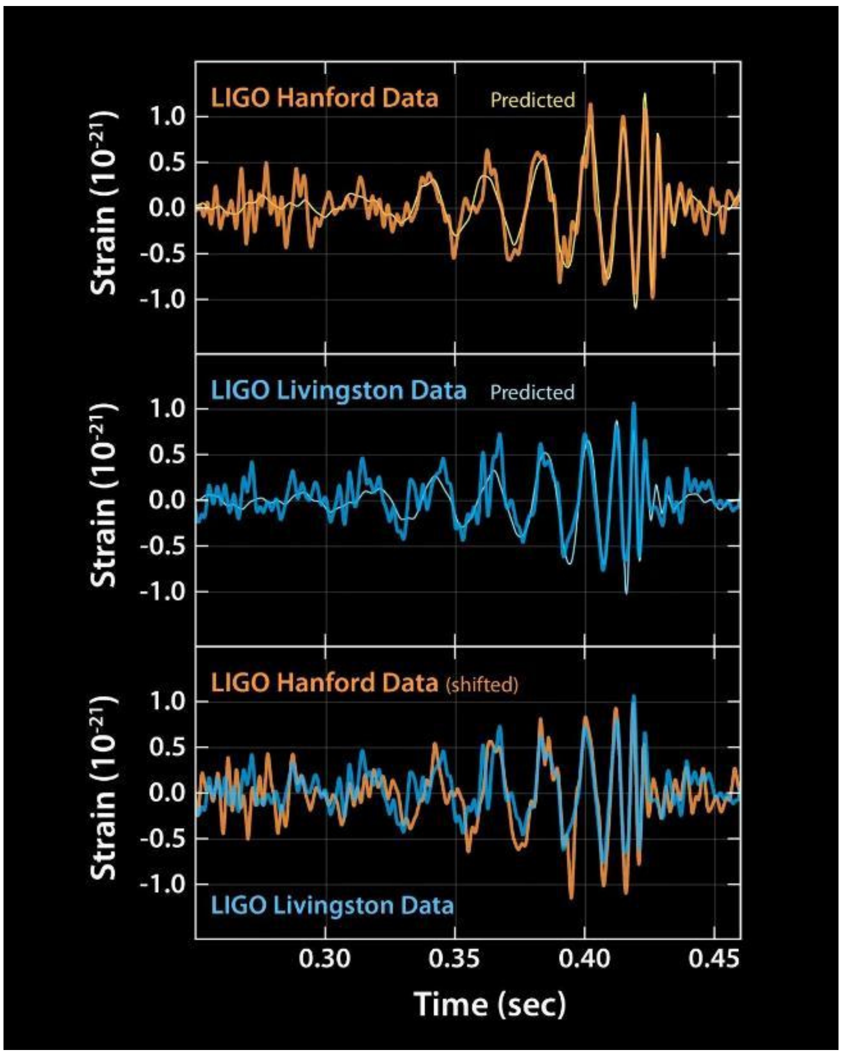


FIGURE 4.5 LIGO gravitational wave data from its two observatories at Livingston, Louisiana, and Hanford, Washington, from the first gravitational wave detection of merging black holes. The lower panels shows the LIGO gravitational wave signal from its two observatories at Livingston, Louisiana (in blue) and Hanford, Washington (in orange, shifted by 7 milliseconds), from the first gravitational wave detection of merging black holes. In the top panel, the signal from the Livingston Observatory is shown with a numerical theoretical model for two merging black holes, each about 30 times the mass of the Sun, lying 1.3 billion light-years away. SOURCE: Caltech/MIT/LIGO Laboratory, https://www.ligo.org/detections/GW150914.php.

Low funding rates at both NASA and NSF have affected the ability to carry out theoretical investigations. For NASA's ATP program, which funds theory relevant to NASA's missions, proposal funding rates dropped from 17 percent in 2010 to 14 percent by 2013.[6] Since 2015, the program has moved to a 2-year proposal cadence, but proposal success rates still remain low (22 percent in FY 2019).[7] When the longer cadence is coupled with low success rates, scientists have little realistic expectation that their research will be funded while it is most relevant, if it is ever to be funded. The NSF AAG program discussed in the section above is a crucial vehicle for funding new independent and novel investigations in all fields of astronomy and astrophysics, but especially in theory. Lack of data on success rates of different proposal types prevents an assessment of how well this program supports theoretical investigations.

Recommendations for augmenting support for theoretical investigations have appeared in multiple previous astronomy decadal surveys, signaling a perennial under-commitment to supporting this mode of scientific inquiry. Most recently, the Astro2010 recommended that funding for NASA's Astrophysics Theory Program be increased by 25 percent, but instead the budget remained flat, and the calls for proposals slowed to a 2-year cadence. When coupled with current extremely low proposal success rates, these changes have particularly hurt the career development of pre-tenure theorists. The Panel on an Enabling Foundation for Research took up the earlier suggestion of a 25 percent increase in funding, and additionally suggested that the ATP be returned to an annual cadence.

New Worlds, New Horizons also recommended the creation of a new inter-agency funding opportunity called Theory and Computation Astrophysics Networks (TCAN). This program was intended to respond to the facts that (1) theoretical and computational problems have reached a scale and complexity that requires more sustained funding of larger teams than a standard NSF AAG or NASA ATP grant; and (2) that many of the most important theoretical problems transcend the artificial boundaries of the three agencies (NSF, NASA, and DOE). The resulting TCAN concept would have supported 5-year programs that would be jointly supported by all three agencies. However, after one funding cycle only NASA has continued to participate in the program.

Finding: A strong foundation of theoretical research remains critical for interpreting astrophysical observations and planning new facilities, but past decadal survey recommendations for supporting theory have not been implemented.

Conclusion: Theoretical investigations are necessary to extract the full scientific intent of new and existing facilities, and funding for such studies must increases to recover from limited funding in the past.

Recommendation: Given the foundational importance of theory to the astronomical enterprise, NASA's Astrophysics Theory Program should resume an annual cadence, and receive a 30 percent funding augmentation in real(inflation-adjusted) dollars over 5 years from 2023-2028 starting with the fiscal year 2019 budget inflated appropriately.

4.2.3 Maximizing the Science through Large Programs Utilizing Ground-Based Facilities.

The method of funding data analysis for NSF-supported projects is significantly different, and at times more problematic, than the funding for NASA projects. Observing time on a NASA facility is accompanied by funding to support data analysis and the creation of high-level data products. This ensures that observers have the resources they need to accomplish the proposed science. For example, the

[6] E. Scannapieco, NASA, presentation to EF panel, and NASA written communication to the steering committee on February 3, 2020.

[7] Ibid.

Hubble Space Telescope mission allocates nearly a third of its annual operations budget to associated research funding; this in turn supports a robust rate of published papers, which peaked in 2019 at 1014 refereed papers that year. The lack of equivalent funding for time awarded on NSF-supported ground-based facilities means that researchers must apply for funding through the AAG proposal process, typically after the observing time is approved. This multi-step process regularly produces delays of two or more years between when observations are approved and when funded scientific personnel can formally join the analysis. The process is also highly inefficient, requiring at minimum two separate rounds of proposals and reviews, once for a telescope time allocation committee and once for a grants panel. Even with higher AAG proposal funding rates, the time delay and gauntlet of multiple proposal reviews add significant inefficiencies that hamper the scientific output of the most powerful facilities.

Finding: Associating research funding for data analysis and production of high level data products with awarding of observing time ensures that observers have the resources they need to accomplish the proposed science.

The NASA model does not translate easily to the ground, however, where weather and observing conditions add significant uncertainty to program completion rates, particularly for smaller observing programs. It would, however, be appropriate for and significantly increase the scientific impact of the subcategory of large projects that exist for current and future MREFC-class astronomical facilities. These MREFC-funded facilities are a significant investment of federal dollars and are motivated by a few major scientific objectives. Large or key projects are the programs that have been established by peer review to be the most important science priorities for a given telescope. They require a large investment in observing time and there is often an expectation of dissemination of results through paper publishing, and catalogs or data releases. These programs are typically given the observing resources to reach a high degree of completion.

There are only a small number of these large programs for MREFC-class astronomical facilities, with typically three to five large programs approved per year on each, with not all programs having significant U.S. participation. Large programs are sufficiently competitive that funding panels are naturally reluctant to award funding to an ambitious proposal that has not yet actually been approved. Thus, projects are not likely to have their challenging processing and analysis funded in a timely manner, which delays the science deemed to be especially important. The net result is that enormous barriers exist for producing the most compelling, legacy science, particularly in the early stages of large projects, when data collection and reduction is most intensive, but funding is 2 or more years out of reach.

Allowing these large programs to immediately submit supporting budgets to the NSF AAG would give U.S. investigators a way to quickly ramp up reduction and analysis, and to focus their energy on producing science rather than yet another proposal to the AAG program. This approach would also help to attract a larger, more diverse user base to NSF facilities, given that lack of funding is a larger barrier for scientists from under-resourced institutions, who are not well-positioned to tackle ambitious projects that would need years of up-front effort before NSF funding can be secured.

Recommendation: The National Science Foundation Division of Astronomical Sciences should establish a mechanism of associated research funding for data analysis and production of high level data products for large principal investigator-led programs on MREFC-scale astronomical facilities in order to accelerate the scientific output and maximize the timeliness and community impact of these key large projects.

Given the small number of large MREFC-scale programs, this recommendation could be accommodated within the AAG increase.

4.3 BREAKING DOWN CROSS-AGENCY BARRIERS

Astrophysical questions increasingly transcend traditional wavelength, division, and agency boundaries. This richness in scientific perspective unfortunately is accompanied by logistical complications in funding and project management. Historical boundaries between organizations can raise a particularly high barrier to fundamentally interdisciplinary efforts, such as solar physics. Within the last decade or so, the emergence of new subject areas has magnified these inconsistencies: the study of exoplanet atmospheres requires knowledge not only of temperatures and pressures appropriate there, but also knowledge gleaned from studies of planets in the Solar System, and indeed detailed understanding of processes at work on Earth; within the framework of NASA's Science Mission Directorate, this endeavor potentially spans three distinct directorates. The rise of multi-messenger astronomy unites information gathered across the electromagnetic spectrum, from ground and space, as well as new carriers like gravitational waves (Figure 4.6) and neutrinos, requiring a breakdown of traditional funding silos to fully realize the science return. These new ways of approaching science can potentially be funded by multiple entities, but simultaneously risk not being funded by any if no group feels they "own" the science. These barriers can be transcended through dedicated programs like the Windows on the Universe (WoU) Initiative (which jointly reviews proposals between NSF's physics and astronomy divisions), or like the TCAN theory program recommended in the 2010 decadal survey (Section 4.2.2). These programs need thoughtful guidelines and execution to be successful in practice. NASA has started taking steps to identify disciplines needing interdivisional research and/or interagency partnerships and coordinating technology development across multiple disciplines.

Foundational cross-agency issues affect fields with new and exciting results like neutrino astrophysics, gravitational wave astronomy, and particle astrophysics. These are relevant to astronomy research but are funded through other divisions. Neutrino astrophysics and gravitational wave astronomy are primarily funded out of NSF PHY and/or Office of Polar Programs. As described in more detail in Chapter 7, the survey committee is endorsing (but not ranking) the IceCube-Gen2 neutrino large facility and technology development for next generation gravitational wave observatories largely because of the benefit to the field of astronomy.

> **Conclusion:** Effective mechanisms to fund cross-cutting research at NSF, NASA, and the DOE would accelerate scientific results.

4.4 SOLAR PHYSICS

Solar physics is directly relevant to astronomy. As the nearest star, the Sun is both a key calibrator for our understanding of stellar astrophysics, and a unique laboratory for understanding magnetism and its coupling to mass, which is relevant through the universe. The Sun is also an important input to understanding Earth's climate and space weather. In the next decade, solar observations and theory will be key ingredients in understanding the Earth-Sun connection and its implications for the co-evolution of stars and planets throughout the Milky Way Galaxy, particularly given the impact that eruptive events and high energy emission of light and particles can have on planetary atmospheres. While the new high-resolution capabilities of DKIST will surely transform our understanding of the Sun, there remains a pressing need for complementary global measurements of the entire Sun and its magnetic activity.

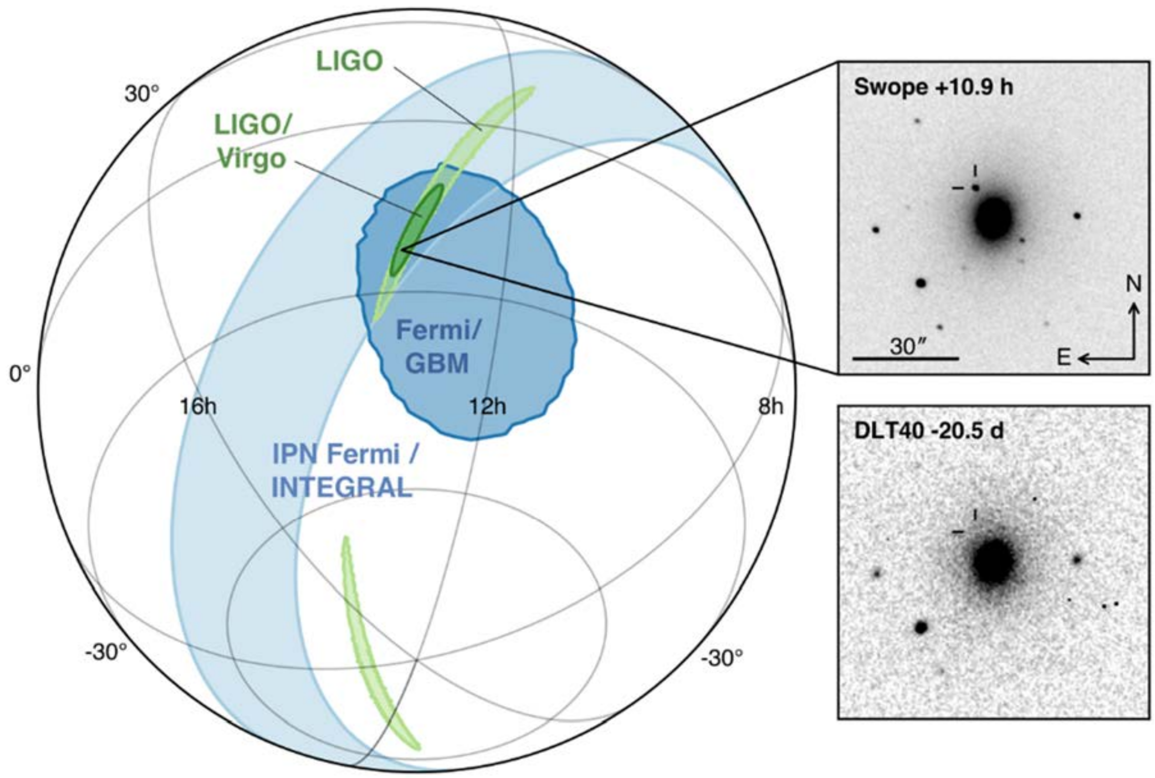


FIGURE 4.6 The multi-messenger nature of the detection of the kilonova 170817, first detected in gravitational waves and gamma ray bursts, and shortly thereafter in many other wavelengths. In the left panel, green contours indicate location determination from gravitational wave detectors (LIGO in light green, LIGO-Virgo in dark green); light blue contours delineate likely regions using triangulation from time delays between gamma-ray satellites Fermi and the INTErnational Gamma-Ray Astrophysics Laboratory (INTEGRAL); and dark blue contours trace the Fermi Gamma-ray Burst Monitor localization. The insets on the right show optical images of the host galaxy NGC 4993 10.9 hours after the merger taken with the Swope telescope (*top right*) and a DLT40 pre-discovery image 20.5 days prior to the merger (*bottom right*). SOURCE: B. P. Abbott et al. 2017, "Multi-messenger Observations of a Binary Neutron Star Merger," *The Astrophysical Journal Letters*, 848 L12. doi:10.3847/2041-8213/aa91c9.

Observations of the Sun depend on facilities spanning multiple federal agencies, even different directorates within the same agency, and these groups take advice from different decadal surveys. Solar ground-based observations are done with a mix of solar-dedicated facilities (such as the Mauna Loa Solar Observatory [MLSO] and the Expanded Owens Valley Solar Array [EOVSA]), as well as general purpose astrophysics facilities like the JVLA and ALMA that have solar-capable instruments. The field of ground-based solar physics is funded by two different NSF divisions: Astronomical Sciences (within the Mathematical and Physical Sciences Directorate) as well as Atmospheric and Geospace Sciences (residing in the Geosciences Directorate). Space-based heliophysics research at NASA is the domain of the Heliophysics Division, distinct from the Astrophysics Division. The National Oceanic and Atmospheric Administration (NOAA) also oversees space weather prediction capabilities and is another federal agency relevant to the subject. The direction for investments in space-based assets are prioritized by the solar and space physics decadal survey process, while this astronomy and astrophysics decadal survey committee advises only the division of Astronomical Sciences at NSF about ground-based solar physics.

This mix of different solar observation regimes, each controlled by a separate decadal process, was a topic of attention in Astro2010, which had recommended that NSF work with several communities to determine the best route to a balanced and effective ground-based solar program maintaining

multidisciplinary ties. In the current decadal process (which can only recommend ground-based components of the solar observation program), only three white papers describing solar facilities were submitted to this decadal survey.[8] These mid-scale ground-based solar projects emerged with favorable reviews from the OIR and RMS panel reports, however the solar mid-scale projects are not considered amongst the strategic initiatives called out in Chapter 7. The survey committee was not able to give proper perspective as to how these facilities will support and enhance the broad range of multi-agency activities currently underway in solar physics, as these are the domain of the solar and space physics decadal survey. Advancing these myriad scientific goals is most efficiently done utilizing a comprehensive approach.

> **Conclusion:** The most appropriate role for future astronomy and astrophysics decadal surveys is to comment on the value of ground-based solar physics projects for astronomy and astrophysics scientific priorities. For consideration of these projects in the context of the full range of multi-agency activities in solar physics, the solar and space physics decadal survey is the more appropriate body to prioritize and rank them.

4.5 THE DATA FOUNDATION

Through much of the past 150 years, the majority of astronomical observations were held by individuals or institutions, archived on photographic plates or data tapes. By 2020 this landscape has been completely transformed. All modern data are digital, and a significant portion are archived in publicly-accessible on-line data libraries. The scientific importance and impact of these archives is fundamental. For the past 15 years, for example, publications from archival use of data from the Hubble Space Telescope have outnumbered those by the original proposing teams (Figure 4.7), with comparable numbers of citations,[9] and other major facilities are seeing similar trends. The empirical evidence it that curating scientific data in well-organized archives enables multiple repurposings and extends the useful lifetime of the data (Figure 4.8).

Astronomy is entering a second wave of this data revolution, with increasing numbers of survey facilities largely dedicated to producing archival data sets from the outset, which are subsequently shared by thousands of users for myriad individual scientific projects. For two decades, the Sloan Digital Sky Survey has been a ground-breaking precursor of this new mode of survey astronomy. In space, NASA's Infrared Astronomical Satellite and Wide-Field Infrared Survey Explorer's all-sky surveys created data sets that have lasting value to this day. More recently, the European Space Agency (ESA) Gaia observatory, which is measuring precise positions and proper motions for a billion stars, has revolutionized Milky Way and stellar astrophysics (Chapter 2). Although it was fully built and supported by Europe, its data archives are openly accessible worldwide, and have supported hundreds of investigations by U.S. astronomers in the 5 years since the first data release. In the coming decade the Vera Rubin and Nancy Grace Roman Observatories, the highest-priority ground and space projects in the 2010 decadal survey, respectively, will provide comparably rich data sets, which promise to revolutionize time domain astronomy and promise breakthrough discoveries across a wide range of astrophysical disciplines. They will also bring unprecedented volumes of data—of order 500 Petabytes (500 million Gigabytes) by the end of 2030 collectively across all observatories and missions, several orders of magnitude more astronomical data than has been collected in human history. When combined with the increasing availability of data from other, mid-scale facilities, the very nature of the observational research enterprise is evolving. In short, progress in astronomy requires fully preparing for the next phase

[8] ngGONG (Hill et al. 2019BAAS…51g..74H), COSMO (McIntosh et al. 2019BAAS...51g.165M), FASR (Bastian et al. 2019astro2020U..56B).

[9] Space Telescope Science Institute, https://archive.stsci.edu/hst/bibliography/pubstat.html, accessed May 18, 2020.

of the on-going transition away from targeted observations to large public data sets, in order to maximize the science returns from current and upcoming facilities.

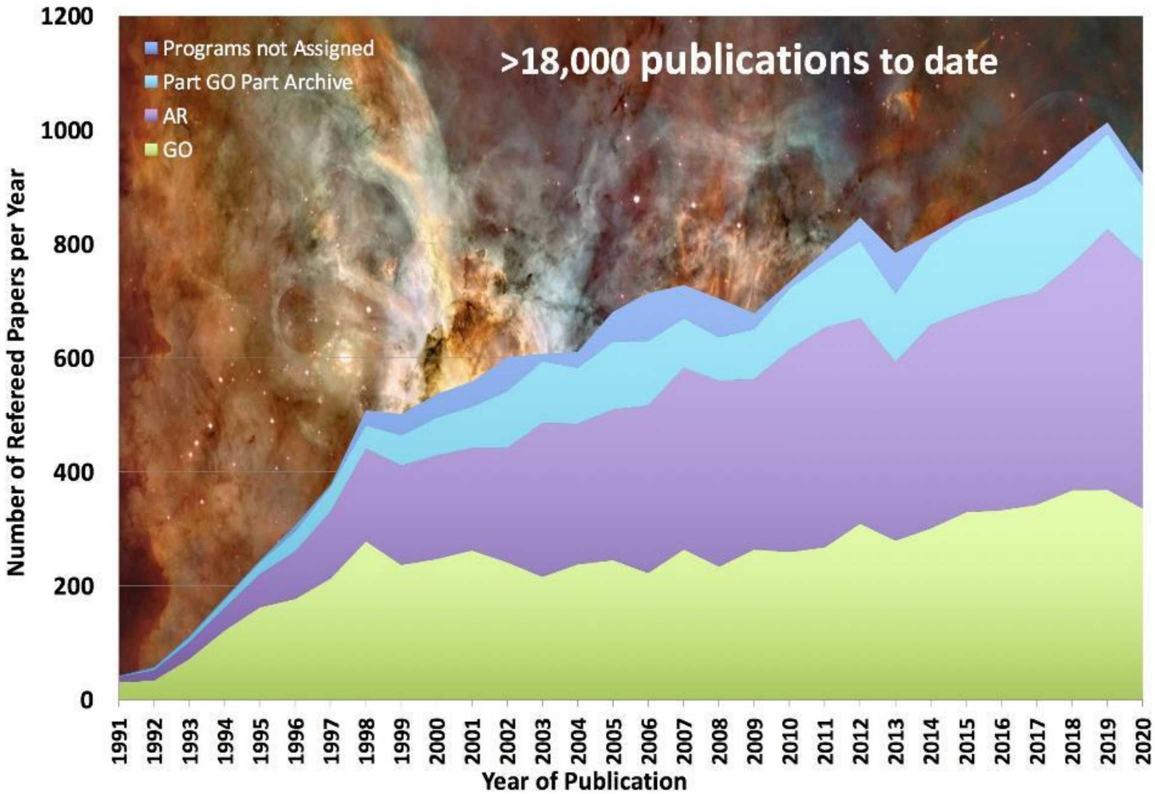


FIGURE 4.7 History of publications arising from observations taken with the Hubble Space Telescope. The green curve gives the trend of refereed papers originating from the original proposing Guest Observer (GO) team; papers in purple have no overlap between the original proposing team and paper authors and indicate purely archival (AR) research uses; the aqua curve indicates a mix of GO and archival researchers. The fourth category indicates papers for which the assignment into the other bins cannot be made. The rate of archival paper production has outpaced that of GO paper production from the early days of the observatory, a product of the open archives and pipeline data processing. SOURCE: R. Osten, based on data available from STScI, https://archive.stsci.edu/hst/bibliography/pubstat.html.

Along with the increasing importance of surveys and large data volumes, a related revolution in computational astrophysics is underway. Numerical simulations are playing an ever-growing role in modeling the physics of planets, stars, interstellar clouds and plasmas, galaxies, and the universe itself. Numerical simulation has become an essential skill set for many theoretical astrophysicists. The outputs from these simulations represent a valuable resource, but currently are rarely made publicly available, and will comprise a very significant data volume. Although many theorists and modelers use publicly available codes, far fewer people write or maintain them.

The data revolution has also transformed the manner in which many astronomers conduct the majority of their research. Many observational astronomers rarely observe in person at a telescope, instead spending the bulk of their time developing methods to carry out sophisticated analyses of large online data sets. The algorithms used to process the data and create the results then become as important as the underlying observations. Mechanisms to share software, such as code-sharing and revisioning through Github and providing tutorials with worked examples through Jupyter notebooks, enable reproducibility and lead to further levelling of the field for access to and improvements on the motivating

science. Directly linking papers and the data contained therein in archives also strengthens the connection between the resultant science and the input observations.

These revolutions are of course not unique to astronomy and astrophysics, and span many fields. "Harnessing the Data Revolution" is one of NSF's "Ten Big Ideas,"[10] initiatives in which NSF plans to make significant investments. NASA has recently convened a "Big Data Task Force"[11] and released the Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024.[12] The recommendations for building the Data Foundation for Astronomy and Astrophysics presented in this section align well with these efforts.

4.5.1 Data Archiving, Curation, and Pipelines

The importance of archiving, curating, and facilitating the use and analysis of these rich data sets has long been recognized by NASA and NSF, and numerous programs exist to support these areas. For NASA these include mission-specific support (Figure 4.8), support for archival data centers, and individual investigator programs such as the Astrophysics Data Analysis Program (ADAP) and Exoplanets Research Program (XRP). NSF supports data curation at its national observatories, and mandates a plan for managing data and sharing the products of funded research in individual investigator programs through its general AAG program. The vast network of private ground-based facilities have much more variable levels of archiving and accessibility. The question is whether the current suite of programs is fit for the 2020's and beyond. While recognizing the many successes of the current programs, the decadal survey will focus on areas where modest investments can produce major scientific payoffs.

Virtually every celestial photon collected by a telescope is a precious resource, capable of contributing to future discoveries. This legacy value can be maximized through investing in infrastructure that enables facilities to collect and reduce these data in a uniform manner, and that archives data to be easily retrievable, with the eventual goal of making the data publicly available. The need for high-level data processing is also being driven by the increasing complexity of instrumentation (e.g., integral-field spectrometers and multi-object spectrographs) in space and on the ground. The importance of joint analysis of observations from different facilities and wavelengths, and of sophisticated archiving with associated science platform tools, will grow dramatically over the next decade. A prime example is the measurement of cosmological constraints on dark energy and other parameters in the coming decade, which will rely heavily on the joint processing and analysis of data from the Euclid (ESA), Roman, and Rubin observatories. As detailed in previous chapters, the tremendous interest in multi-wavelength, multi-messenger, and time-domain analysis will pose new challenges over the next decade, as will carrying out any science project with unprecedented volumes of data.

The current state of these data archives varies considerably, but the general trend is for an increasing role of archival data in scientific pursuits. The decade just completed saw an expansion of archive capabilities both on the ground and in space (Figure 4.7, 4.8, 4.9, 4.10, 4.11), and this is only expected to grow in the coming decade. The remote nature of space facilities mandated effective data storage from the outset, and perhaps not surprisingly, well-managed archives are available for nearly all major NASA missions. These data have a long duration impact: data taken from the early days of the Hubble Space Telescope still find productive uses in refereed papers nearly 30 years after initial acquisition. Seventy percent of data archived from early in the life of the Chandra X-ray Observatory appear in four or more publications (Figure 4.8).

The situation for ground-based facilities is much more mixed. Large facilities such as those built and operated by the International ALMA Observatory and the European Southern Observatory (ESO), as well as surveys such as SDSS and the Panoramic Survey Telescope and Rapid Response System (Pan-

[10] https://www.nsf.gov/news/special_reports/big_ideas/index.jsp.
[11] https://science.nasa.gov/science-committee/subcommittees/big-data-task-force.
[12] https://science.nasa.gov/science-red/s3fs-public/atoms/files/SDMWG%20Strategy_Final.pdf.

STARRS) have established archives of quality rivaling those of the space observatories. These archives have been major contributors to the scientific productivity of those endeavors.[13] Figure 4.9 and 4.10 detail the steady increase in archival usage of ESO Paranal telescopes and the ALMA Observatory, respectively; in both cases roughly a third of papers are now produced using at least some archival data. The effort put in by the ALMA Observatory to create data reduction and calibration pipelines has the result that currently 95 percent of data is calibrated and imaged;[14] in addition to enhancing the archival utility of the data, such steps reduce barriers to entry for new users and widen ALMA's user base. NOIRLab hosts an Astro Data Lab as a centralized hub for archiving and disseminating observations from U.S. nighttime OIR observatories, with emphasis on large surveys and data discovery tools, and the NASA-funded Keck Observatory Archive curates Keck data at the NASA Exoplanet Science Institute (NexScI). Such examples, however, have been the exception rather than the rule. Several factors account for this situation. Few privately-supported U.S. ground-based observatories are financially positioned or structurally incentivized to provide fully-reduced data products, and for older public facilities like the JVLA or VLBA with complicated data processing, such a goal may simply not be possible for all data in spite of best efforts (see Figure 4.11 for the increasing trend in archival usage from the JVLA). And while some facilities place their data into public archives, these resources are often difficult to tap. The net result is an opportunity lost, for the scientists who could be exploring data immediately rather than spending months reducing it or making new observations, for the observatories that invested in instruments whose data are underused, and for the science that could be done if that data could be easily accessed.

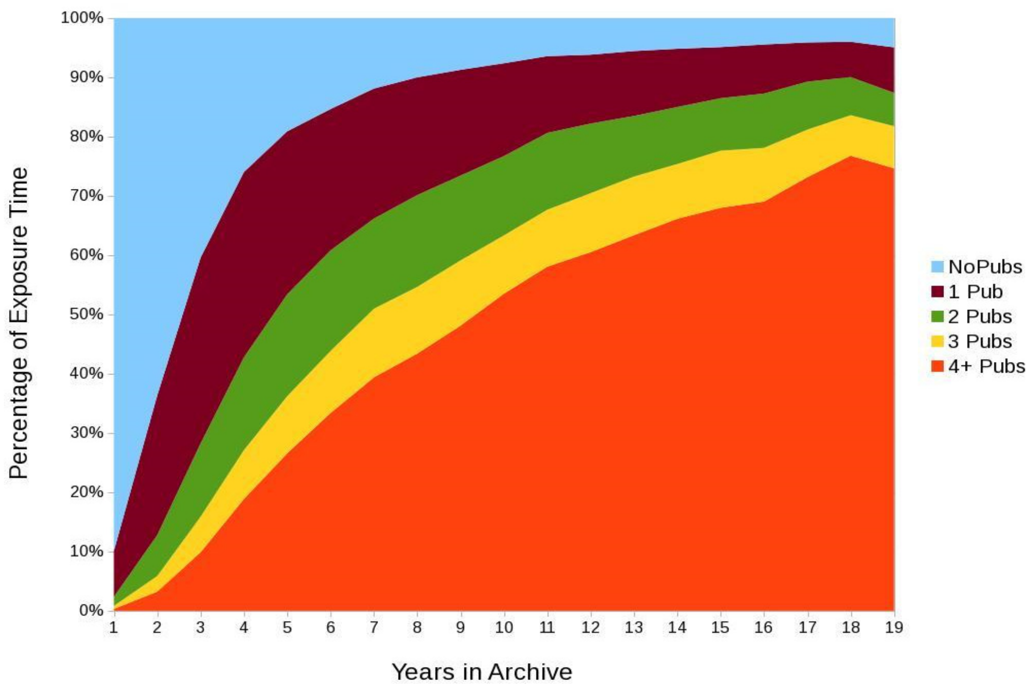


FIGURE 4.8 The percentage of data published as a function of time for data taken from the Chandra X-ray Observatory archive, demonstrating the impact of a well-organized archive. Data here is quantified as exposure time. Here, 70 percent of the oldest data sets have four or more publications using the data. SOURCE: Courtesy of the Chandra Data Archive operations team.

---

[13] M. Romaniello, M. Arnaboldi, C. Da Rocha, C. De Breuck, N. Delmotte, A. Dobrzycki, N. Fourniol, W. Freudling et al., 2016, The growth of the user community of the La Silla Paranal Observatory science archive, *The Messenger*, 163(5), http://www.eso.org/sci/publications/messenger/archive/no.163-mar16/messenger-no163-5-9.pdf.

[14] ALMA, "Processing," accessed on May 18, 2021 https://almascience.nrao.edu/processing/science-pipeline.

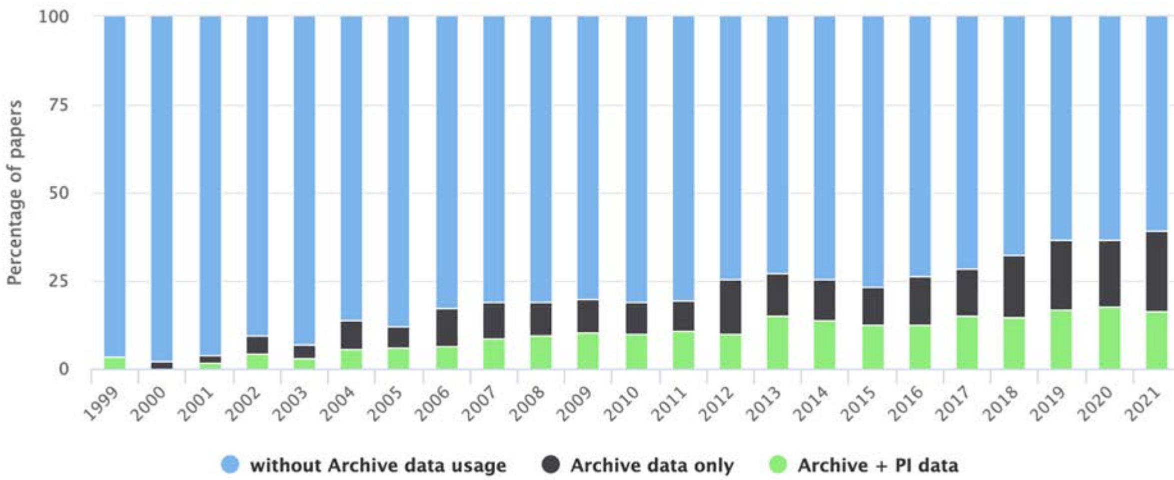


FIGURE 4.9 Graph of the growth of refereed papers using archival data from European Southern Observatory's (ESO) La Silla Paranal Observatory, as a percentage of the total number of papers published that year. Blue indicates papers for which there is overlap between the paper authors and the proposing PI and Co-Is. Black indicates no overlap (i.e. purely archival usage), and green is intermediate, where a combination of purely archival and purely PI data is used. In the last complete year for which statistics are available (2020), more than a third of all papers used archival data in some format. SOURCE: Retrieved from ESO, http://telbib.eso.org/index.php?boolany=or&boolaut=or&boolti=or&yearfrom=1996&yearto=2021&boolins=or&booltel=or&site=Paranal&search=Search. Courtesy of the ESO Telescope Bibliography (telbib), maintained by the ESO Library.

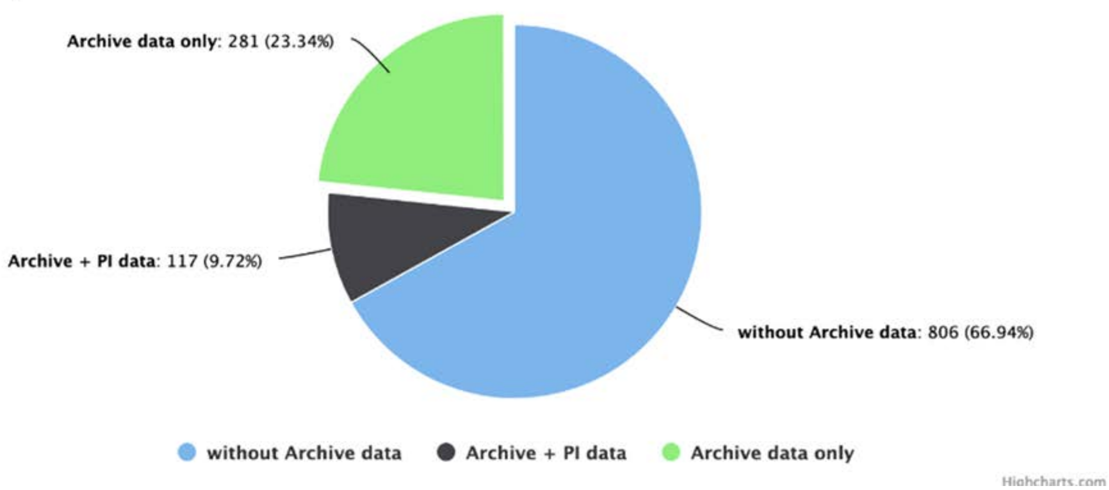


FIGURE 4.10 Archive data usage for refereed papers reporting science results from the Atacama Large Millimeter/submillimeter Array (ALMA) over the 2010-2021 time period. Roughly one third of all papers produced have utilized archival data, either alone or in combination with PI data. SOURCE: ESO/ALMA, http://telbib.eso.org/statistics/archive.php?boolany=or&boolaut=or&boolti=or&yearfrom=2010&yearto=2021&boolins=or&telescope[]=%22ALMA%22&booltel=or&site=Chajnantor&fl=telescope,datastatus&stats=arc&query_stats=year%3A%5B2010+TO+2021%5D+and+site%3AChajnantor+and+%28telescope%3A%22ALMA%22%29. Courtesy of the ESO Telescope Bibliography (telbib), maintained by the ESO Library.

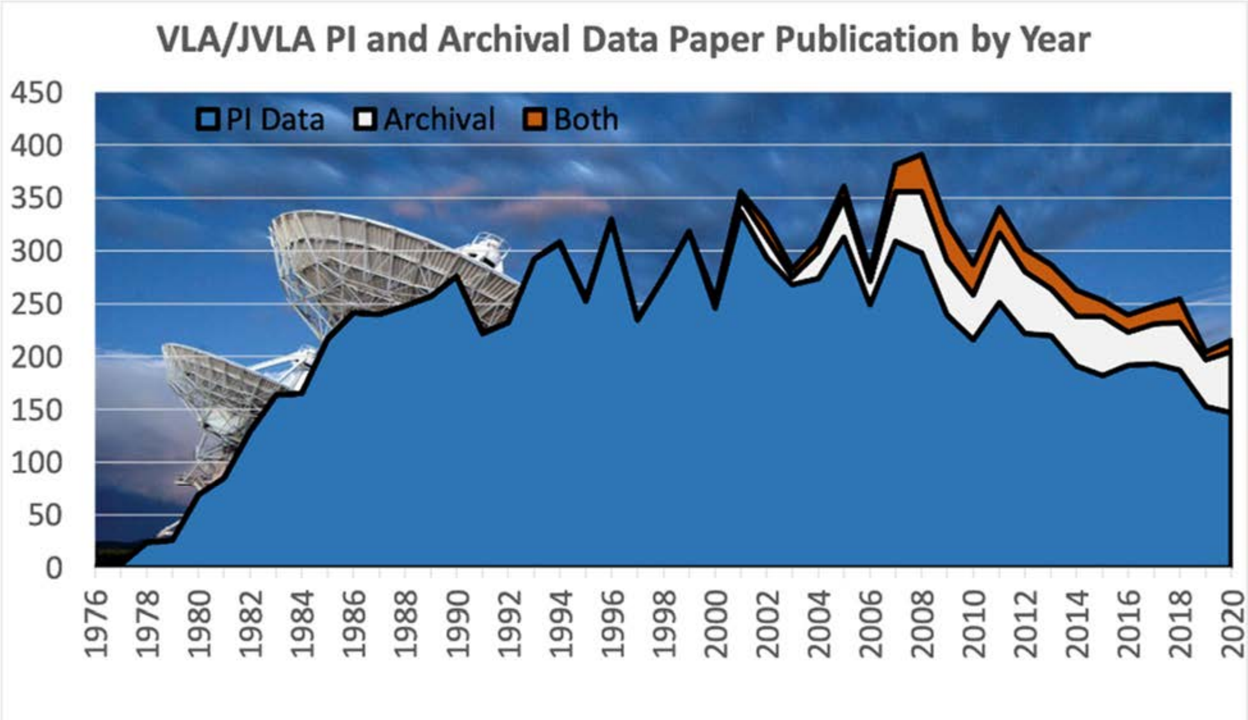


FIGURE 4.11 Paper production from NRAO's Very Large Array/Jansky Very Large Array detailing the evolution of PI usage and archival usage of VLA/JVLA data in producing refereed papers. This does not include papers from surveys such as the NRAO VLA Sky Survey (NVSS) or Very Large Array Sky Survey (VLASS). SOURCES: (*background image)*, https://public.nrao.edu/news/the-very-large-array-astronomical-shapeshifter/. (*main image)*, R. Osten, based on data from L. Utley, NRAO.

This untapped collection of observations not yet being archived can be seen as a tremendous opportunity to extract more science from the U.S. ground-based system. With appropriate strategic planning and modest financial investments, creating and archiving science-ready data products should offer a multi-fold return on the science from ground-based facilities. Scientists' limited time and grant support could be focused entirely on analysis and discovery, and the impact of a telescope's observations could span decades, as photons are reused for science that was unimagined at the time they were collected. Increasing access and the quality of archival observations can also serve as a powerful agent towards broadening participation in the profession, because they bring cutting edge data to any individual with internet access (even the public via Citizen Science initiatives), with minimal barriers to entering the active research community. This democratization of science through archive access will continue in the next decade.[15]

> **Finding:** As demonstrated by space missions, and supported by archiving efforts at ESO and ALMA, readily accessible data in both raw and reduced form from ground-based telescopes can greatly multiply their scientific impact, even more so if pipelines are available to produce processed data.

[15] J. E. G. Peek, V. Desai, R. L. White, R. D'Abrusco, J. M. Mazzarella, C. Grant, J. L. Novacescu, et al., 2019, Robust archives maximize scientific accessibility, white paper submitted to the Astro2020 decadal survey, https://arxiv.org/abs/1907.06234.

With the emerging roles of mega-survey facilities such as the Vera Rubin Observatory, multi-wavelength and multi-messenger astrophysics, and time domain astronomy in this decade, there is an even greater need for data discovery and analysis across multiple archives; this motivates the need for coordination of those archives. The Enabling Foundations panel and this committee considered how to best address this need. The panel proposed establishing a cross-agency umbrella organization called the Astronomical Data Archiving System (ADAS) to coordinate the activities of the existing astronomical data centers and set priorities for new investments. The National Virtual Observatory (NVO) effort was undertaken from 2007-2014 with similar goals of enhancing archive interoperability, but with a significantly different structure and implementation from the proposed ADAS. In partnership with a similar International Virtual Observatory organization, the NVO achieved a number of important successes including the creation of a common set of data formatting and metadata standards, a first generation of data retrieval and exploration tools, and enhancing communication between the many individual data centers in the United States and around the world. NSF support ended in 2014, however, and any new organization of this type would need to build on the lessons learned from the NVO experience, and the experiences of the many archiving centers that have been operating independently over the last two decades. It will be important to preserve the expertise and resources of the existing data centers, in part by providing flexible and stable career paths for archive scientists and software developers. It is also critical to provide a centralized channel for input from the U.S. astronomy community on prioritization of data archiving efforts, so that the work of the archive centers remains in touch with the science needs of the community. The system as envisioned by the Enabling Foundations panel could also address cross-agency strategic planning in the related areas of software development, high-performance/high-throughput computing, archiving and curating data from theoretical simulations, and community training in related areas.

An important component of creating effective archives is coordinating with cross-agency and international archiving services to develop best practices and interoperability. While the International Virtual Observatory Alliance continues, the lack of a U.S. national coordinating effort hampers efficient communication between the various national and international funding agencies and institutes which produce and archive astronomical data. Progress will come from an end-to-end approach that considers the entire flow of data from the instrument, to the archive, to analysis and publication. Increasing the prevalence of both science-ready data products and effective archives is best achieved if done hand-in-hand with each other. Making codes publicly available will help to minimize redundancy, encourage the adoption of common standards, and promote applications using multiple data sets. The survey committee endorses the importance of the goals of the proposed ADAS articulated by the Enabling Foundations Panel, but concluded that the appropriate scope of this effort and the details of the form it would take require further study, led by NASA and NSF, with possible participation from DOE.

Recommendation: NASA and the National Science Foundation should explore mechanisms to improve coordination among U.S. archive centers and to create a centralized nexus for interacting with the international archive communities. The goals of this effort should be informed by the broad scientific needs of the astronomical community.

The U.S. ground-based OIR system is distributed amongst public and private facilities and therefore needs special consideration. For three decades many of the needs in this area were served by the Image Reduction and Analysis Facility (IRAF), a freely-available workhorse software system funded by multiple federal streams.[16] Originally developed and maintained by the National Optical and Infrared Observatories (NOAO) in the 1980's, lack of funding for modernizing the software resulted in its evolution to a community-supported platform on GitHub. This ground-up community-based approach has since become the dominant mode for providing common data pipelines and analysis tools across ground-

[16] Image Reduction and Analysis Facility, "Community Distribution," accessed on May 18, 2021 https://iraf-community.github.io/.

based OIR astronomy. Prime examples include The Python Spectroscopic Reduction Pipeline (Pypelt) which has been adopted for multiple telescopes and instruments for at least 11 observatories,[17] and the broad suite of application Python-based software collected by the Astropy project.[18]

Although these community based efforts have done much to fill the gaping needs for up-to-date pipelines and software, some may need to abandon their support, in the same way the IRAF project was eventually forced to do, because of the lack of sufficient and reliable funding and a stable workforce of contributors. These examples are illustrative of how the normal grant funding structure works poorly in the context of building software infrastructure intended to undergird most modern astronomical software. Any effort to create this type of infrastructure relies on continuous fundraising efforts with short time horizons, without any path to earning the longer-term commitments that are necessary for longer term planning and stability. NSF could help to provide foundational support for these efforts, for example by incentivizing the developments of pipelines and archiving by requiring a reduction and distribution plan as part of its funding for instrumentation, and/or by being open to joint proposals by multiple investigators or observatories to fund adapting and running existing pipelines. Funding for tailoring and operating these pipelines could be granted to observatories in exchange for their willingness to distribute their raw data and science-ready data products to a public archive; the cost of this investment is a small fraction of all the money invested in these facilities already, offering a highly-leveraged scientific opportunity.

Recommendation: The National Science Foundation and stakeholders should develop a plan to address how to design, build, deploy, and sustain pipelines for producing science-ready data across all general-purpose ground-based observatories (both federally and privately funded), providing funding in exchange for ensuring that all pipelined observations are archived in a standard format for eventual public use.

4.5.2 Software Development

Astronomy has entered an era in which well-designed and well-constructed software can be as important for the success of a project as hardware. Examples of highly complex software include pipelines that reduce data for telescopes (e.g., Astropy), data analysis packages, and codes that simulate physical processes, such as stellar evolution (e.g., Modules for Experiments in Stellar Astrophysics [MESA]), N-body (Galaxies with Dark Matter and Gas Interact [GADGET]), or hydrodynamics codes (e.g., Enzo). In addition, advanced statistical techniques and Machine Learning are playing a growing role in reducing large data sets in physics and astronomy, and can also require complex codes. Increasingly, many software packages are developed by large teams, and must make use of heterogeneous types of hardware platforms, from general purpose CPU's running on laptops to large multi-core computing clusters that make use of massive parallelization and graphical processing units (GPUs).

Despite the increasing importance of software development and developers for the advancement of the field, neither are sufficiently funded or supported by existing structures. Moreover, people who have strong software development skills are critical for the field, yet are likely to have many career opportunities outside of astronomy. Indeed, this is true throughout the physical sciences. Professional tracks available for scientists who choose to specialize in developing scientific software infrastructure, which might not be readily supported through traditional tenure-track faculty positions, could aid in the retention and development of these individuals. These positions can be supported through national labs, science centers, observatories, or in research positions at universities, with the understanding from

[17] J. X. Prochaska, J. F. Hennawi, K. B. Westfall, R. J. Cooke, F. Wang, T. Hsyu, F. B. Davies et al., last update May 19, 2020, "PyPelt: The Python Spectroscopic Data Reduction Pipeline," arXiv.2005.06505v2.

[18] The Astropy Project, "Homepage," accessed on May 19, 2021, https://www.astropy.org.

funding agencies that proposals for funding software infrastructure may look more like “instrumentation” proposals than standard PI grants. As discussed in the Open Source Software Policy Options for NASA Earth and Space Sciences report, funding for software maintenance and for open-source software projects, which have been transformative for astronomical science over the past decade, could pay major dividends in the future. [19]

> **Finding:** Software development has become an essential part of every sub-field of astronomy. However, software developers and large software development efforts are not adequately funded or supported by existing structures.

4.5.3 High Performance and High Throughput Computing

Computation has assumed an increasingly pervasive role throughout astronomy and astrophysics, from theoretical simulations of physical processes to sophisticated data analysis. Access to and expertise in the use of specialized computing facilities has therefore become ever more integral to the scientific process, and thus requires on-going investments and training over the coming decade. High-performance and high-throughput computing resources (HPC and HTC, respectively) are playing an increasingly important role in astrophysical research (Figure 4.12), with the former being critical for simulations and the latter for analysis of large data sets. HPC is a major part of a computational astronomer’s climate footprint, hence motivating the use of efficient options. Industry-provided options for HTC currently exist through cloud computing, and are often cost effective solutions to astronomical needs. However, as the size of data analysis problems expand, the cost trades could potentially become unfavorable, and a publicly-funded alternative may be more cost effective than relying on private industry to provide cloud computing. Funding programs may need to adapt to this rapidly changing trade space, while also ensuring that mechanisms exist for proposals to fund cloud computing access, rather than more traditional purchases of computing hardware.

DOE and NSF have announced plans to significantly expand their HPC/HTC capabilities over the coming decade, while NASA plans a more modest expansion. NSF computing resources are also available without NSF support, but this is not true for NASA computing resources. Just as Section 4.3 emphasized the need for inter-agency funding opportunities to support science that transcends agency boundaries, it is essential for agencies to provide opportunities for access to HPC/HTC computing resources for cross-cutting projects. Developing the specialized codes that are competitive for allocations on large national computing facilities requires expertise in both computer science and astrophysics, as well as pre-existing access to facilities that can be used for code development and testing. These requirements can pose a significant barrier to entry for scientists at institutions that do not have access to this expertise or these facilities. Support for training (from NSF, NASA, or national laboratories) can be an effective means of helping to level this playing field. This support could take many forms, for example through small grants to individuals, support for training workshops or schools, or training opportunities through NSF and/or NASA centers.

[19] National Academies of Sciences, Engineering, and Medicine, 2018, *Open Source Software Policy Options for NASA Earth and Space Sciences*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/25217.

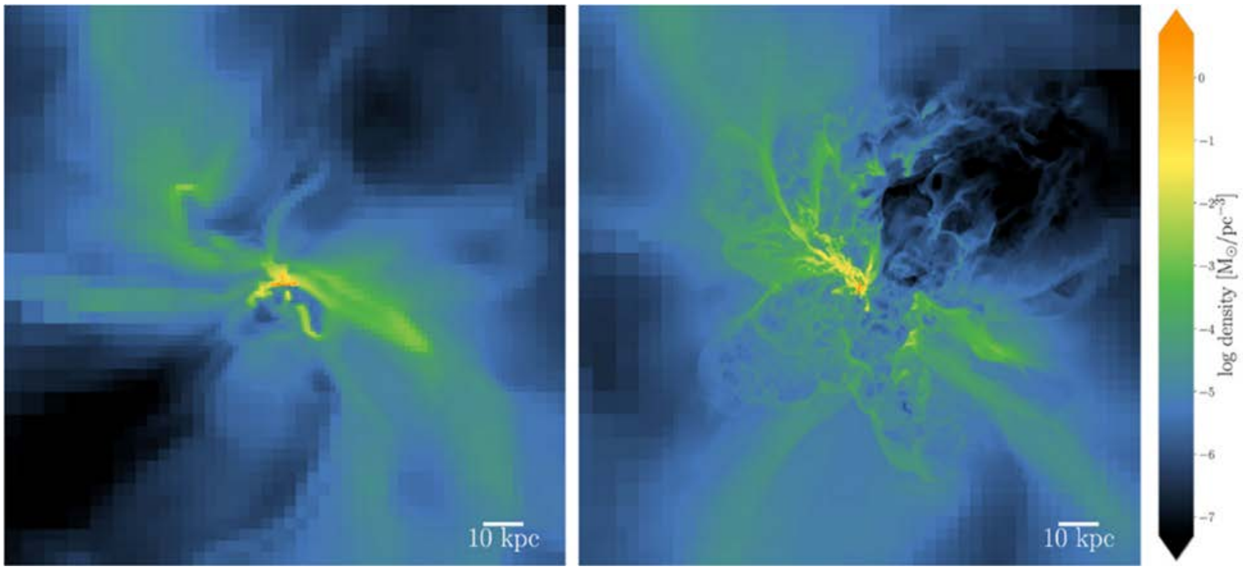


FIGURE 4.12 Comparison of density through the central halo of a galaxy in a standard resolution (*left*) and high resolution (*right*) simulation at a redshift where the majority of star formation is occurring, demonstrates the utility of high performance computing simulations for advancing understanding of complex processes like the factors affecting galaxy formation and evolution. SOURCE: Adapted from Molly S. Peeples et al 2019, "Figuring Out Gas & Galaxies in Enzo (FOGGIE). I. Resolving Simulated Circumgalactic Absorption at $2 \leq z \leq 2.5$," *The Astrophysical Journal*, 873 129. © AAS. Reproduced with permission. doi:10.3847/1538-4357/ab0654.

4.5.4 Data Science and Machine Learning

Over the past decade, data science has advanced dramatically. Machine learning techniques are playing an increasingly important role in astrophysics and this trend is likely to continue into the future. Over the past few years, universities have created multiple joint data science/astrophysics faculty appointments, and are adding new courses. Both undergraduates and graduate students are pursuing joint degrees in programs that did not exist in 2010, and NSF is increasingly investing in "big data" across all subfields.

Astronomical data offer many opportunities for data science research. For example, a key paper in the Data-Driven Discovery Initiative by the Moore Foundation ranked the SDSS as the 6th most influential work in data-driven discovery, just behind Shannon's classic information theory.[20] Astronomical data are valuable for data science for many reasons: the data sets are rich and openly available, well-structured and well-modelled. These have led to numerous new techniques for likelihood-free inference, advances in density estimation, implicit generative models, and probabilistic programming. These techniques are now being used across a wide range of fields (e.g., particle physics, chemistry, and neuroscience) and are part of an emerging new area spanning machine learning and the physical sciences.

Data science offers powerful new tools for studying astronomical data and astrophysical systems. Machine learning has already shown significant success at providing tools for identifying anomalies in data, and can speed up parameter estimation in large data sets by significant factors (Figure 4.13). These techniques could lead to transformative discoveries from the new data sets available in the 2020s. Machine learning has the potential to increase the amount of information obtained from astronomical data sets by enabling modeling of complex non-linear phenomena and instrumental effects. If it can be

[20] M. Stalzer, C. Mentzel, 2016, A preliminary review of influential works in data-driven discovery, *SpringerPlus*, 5:1266, https://doi.org/10.1186/s40064-016-2888-8.

successfully used to model multi-scale phenomena, it could open up the ability to more accurately simulate a wide range of astronomical processes from planet formation to galaxy formation.

> **Finding:** Data science, including applications of machine learning, will play an increasing role in astronomical research over the coming decade. Incorporating training in this area at the graduate level and beyond will better prepare researchers regardless of whether they pursue careers in astrophysics or in other STEM fields.

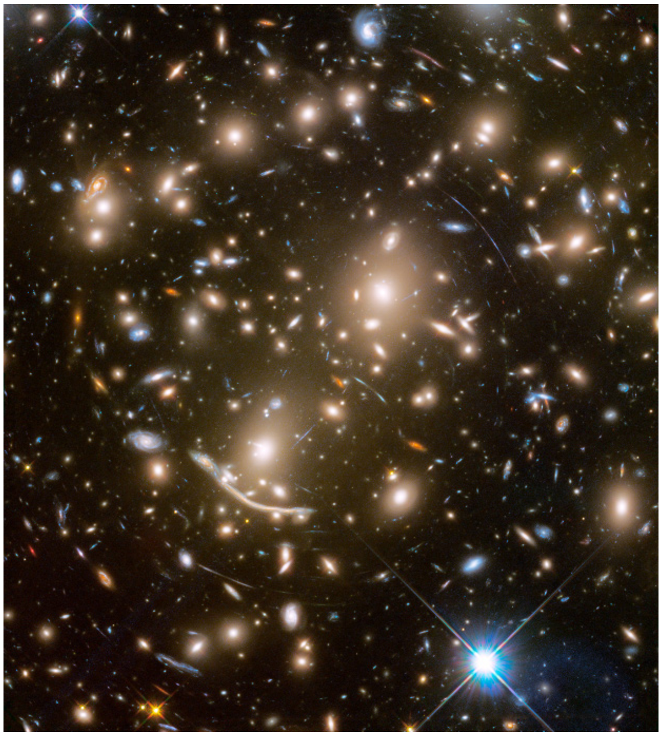


FIGURE 4.13 Hubble Frontier Fields image of the galaxy cluster Abell 370, illustrating numerous arcs resulting from strong gravitational lensing of background galaxies as their light passes through the massive cluster and is subsequently distorted. Machine learning has demonstrated an ability to identify strong lensing arcs orders of magnitude faster than the current state of the art (references in Ntampaka et al. 2019)[21] and is an example of the impact of using deep learning techniques for model parameter estimation in large data sets. SOURCE: Space Telescope Science Institute, https://frontierfields.org/. NASA, ESA, and J. Lotz and the HFF Team (STScI).

4.5.5 Laboratory Astrophysics

At its core, the science of "astrophysics" is built around the assumption that the observational data which astronomers and astrophysicists collect are all produced by understandable, physical processes that are the same throughout the universe. As observations are pushed to rarer or fainter spectral features, or to previously unexplored systems or physical conditions, the understanding is being increasingly limited not by the quality of the data themselves, but by the limited information about the underlying physical parameters.

Thankfully, many of the needed parameters can be measured here on Earth. Laboratory measurements are needed to determine the oscillator strengths of atomic, ionic, and molecular transitions;

[21] M. Ntampaka, C. Avestruz1, S. Boada, J. Caldeira, J. Cisewski-Kehe, R. Di Stefano, C. Dvorkin, et al., 2019, The role of machine learning in the next decade of cosmology, white paper submitted to the Astro2020 decadal survey, https://arxiv.org/pdf/1902.10159.pdf.

reaction rates for the interactions that control the abundances of astrophysically relevant gases, ices, solids, and high-energy tracers; and the complex surface chemistry and optical emission and scattering processes that are increasingly relevant for understanding solid material in the ISM, protoplanetary disks, planetary atmospheres and planetary and stellar interiors. These experiments can be challenging, since conditions found in astronomical settings span a wide range of conditions that can be difficult to match in Earth-based laboratories.

Spectral surveys in the far-infrared and submillimeter with Herschel and ALMA have detected more than 100,000 spectral lines from molecular clouds, star forming regions, and the center of the Milky Way (see Box 4.1). Many of these lines are from a few abundant molecules, such as methanol. However, even these spectral "weeds" contain useful information on conditions in the interstellar medium, such as temperatures, densities, pressures and radiation fields. Molecules of potential prebiotic interest, such as amino acids, have complex spectra requiring detections of multiple lines with a common excitation. Infrared spectra from JWST will have spectral features from aromatic hydrocarbons, which contain information on radiation fields and annealing processes. Studies of exoplanet atmospheres will need chemical reaction rates for a broad array of chemical species and conditions. Understanding the chemistry of protoplanetary disks is the first step toward grasping the composition and evolution of planetary systems. More generally, interpreting these spectral lines requires advances in laboratory astrophysics.

Despite limited resources, examples abound of areas in which laboratory astrophysics has been key to advancing astrophysical discoveries over the past decade (Box. 4.1). In the search for humanity's interstellar chemical origins, the 2010s delivered first identifications of aromatic organics, and chiral molecules, and the first inventories of organic molecules at the onset of planet formation. These results were obtained because of new spectroscopic line lists. Complementary laboratory work revealed that many of these organics can form in icy grain mantles at close to absolute zero temperature, and that complex, prebiotically interesting organic molecules are thought to be ubiquitous during star and planet formation. New laboratory data has also been key to characterize the atmospheres of exoplanets; experimentally determined molecular line opacities at high temperatures have enabled retrievals of water abundances and constraints on atmospheric carbon/oxygen ratios, while haze formation experiments have been key to elucidate what kind of hazes and clouds may form on different kinds of exoplanets. Laboratory astrophysics has also been instrumental in advancing the fundamental understanding of the underlying physics governing stars, for example in significantly revising constraints on convective mixing in the Sun and other stars. If astronomy aims to understand the structure and evolution of stars, galaxies, and the universe as a whole through observations from future facilities, laboratory astrophysics will be required.

BOX 4.1 Applications of Laboratory Astrophysics

The history of buckminsterfullerene, or "buckyballs," illustrates the interdependence of theory, laboratory work, and astronomical observations. One of the theoretical motivations that led to the discovery of these soccer ball-shaped carbon molecules was a desire to understand the diffuse interstellar bands. The origins of these broad absorption features in astronomical spectra remained elusive for nearly a century, although large carbon molecules in interstellar gas clouds were considered likely candidates. Laboratory experiments in the 1980s led to the identification of C_{60} in the emission spectra of soot. The solid phase of C_{60} in astronomical spectra was first identified in the mid-infrared spectrum of a star in 2012. Since then, absorption spectra of heavily reddened stars have revealed multiple transitions of singly ionized C_{60}, confirming its presence as one of the carriers of the diffuse interstellar bands (Figure 4.1.1).

The spectral signature of fullerenes could not have been identified without theoretical and laboratory studies of soot. In the era of ALMA, JWST, and proposed future facilities, with which astronomers will have the capability to study the chemical origins of exoplanetary systems and to detect molecules in exoplanetary atmospheres, the ability to spectroscopically identify complex

chemical species in space, including prebiotic molecules, is critically important. A robust program of laboratory astrophysics to support these investigations is essential.

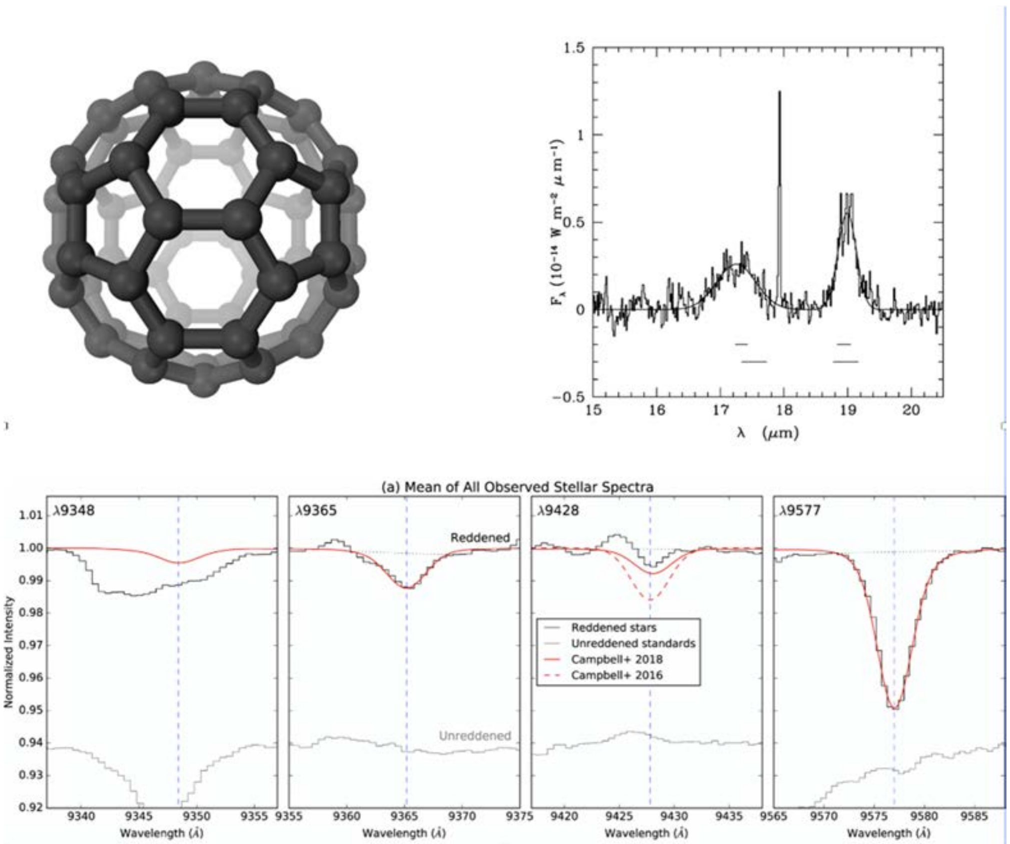


FIGURE 4.1.1 (*upper left*) Image of buckminsterfullerene, C_{60}, indicating the linked carbon atom structure. (*upper right*) Spitzer spectrum of possible C_{60} features in the infrared spectrum toward the binary XX Oph. This is the first detection of C_{60} in the solid phase. Horizontal lines indicate the location of features in C_{60} smoke and gaseous C_{60}. (*lower panels*) Absorption spectra from Cordiner et al. (2019) with the Hubble Space Telescope, comparing those of reddened and unreddened stars. Red curves denote broadened laboratory spectra of transitions of C_{60}^{+}. SOURCE: *Upper left:* Buckyball graphic in the public domain, from Benjah-bmm27, https://commons.wikimedia.org/wiki/File:Buckminsterfullerene-perspective-3D-balls.png. *Upper right:* A. Evans, J.Th. van Loon, C.E. Woodward, R.D. Gehrz, G.C. Clayton, L.A. Helton, M.T. Rushton, S.P.S. Eyres, J. Krautter, S. Starrfield, and R.M. Wagner, 2012, Solid-phase C_{60} in the peculiar binary XX Oph?, *Monthly Notices of the Royal Astronomical Society: Letters* 421(1): L92-L96, doi: 10.1111/j.1745-3933.2012.01213.x, by permission of the Royal Astronomical Society. *Lower panels:* Adapted from M.A. Cordiner et al., 2019, Confirming interstellar C60+ using the Hubble Space Telescope, *Astrophysical Journal Letters* 875: L28, doi:10.3847/2041-8213/ab14e5, © AAS, reproduced with permission.

The 2020s will also see an even greater focus on stellar astrophysics, with “industrial scale spectroscopy” combining with data from Gaia aimed at obtaining complete inventories of stellar properties, such as detailed chemical compositions, masses, and ages. In the era of upcoming photometric (Vera Rubin Observatory, Skymapper,[22] etc.), and large high- to low-resolution spectroscopic surveys (SDSS-IV, SDSS-V, the 4-metre Multi-Object Spectrograph Telescope [4MOST], the William Herschel Telescope Enhanced Area Velocity Explorer [WEAVE], Galactic Archaeology with

[22] Keller, S. C. et al., 2007, Publications of the Astronomical Society of Australia, 24, 1.

HERMES [GALAH], Gaia-ESO, etc.) astronomers will not be limited by data in the pursuit of stellar astrophysics, but rather by a lack of laboratory measurements needed to interpret the data. While these fundamental parameters are crucial for stellar astrophysics, they are also important in a wide range of astrophysics ranging from exoplanet science to galaxy formation. The availability of relevant laboratory atomic, molecular, and optical (AMO) data, such as highly accurate wavelengths, transition probabilities, photoionization cross sections, line broadening parameters, and collisional cross sections, will be critical for maximizing the scientific return of these surveys, observatories, and missions, which together represent a significant investment of U.S. astronomy resources. At higher energies, the scientific return from high-resolution X-ray spectroscopic missions, such as XRISM and Athena, will not be able to capitalize on their high-resolution capabilities without new atomic data including collisional and photoionization cross sections and dielectronic recombination rates. Potential diagnostics of density, temperature, ionization, abundances etc. will not be realized without improved laboratory data on transition energies, electron impact ionization collision strengths, photoexcitation, and ionization. Laboratory astrophysics is also a required foundation to enable science on a range of scales—from as small as dust grain growth to the solar convection boundary problem, to understanding the shock physics of supernovae. A prime topic for the next decade, constraining the heavy elements produced in the electromagnetic counterparts to neutron star mergers (kilonovae) requires understanding the spectra of rapid neutron capture heavy elements such as neodymium and other rare-Earth metals (Figure 4.14). There are insufficient laboratory measurements of line strengths and wavelengths for these elements so current models that predict and interpret observations rely on theoretical atomic structure calculations. Additional laboratory measurements would be very valuable and would also inform abundance measurements of neutron rich elements in stellar spectra.

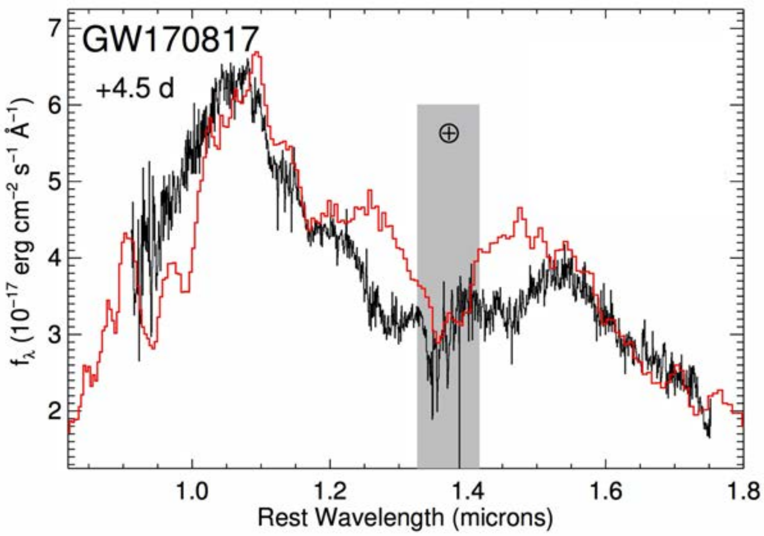


FIGURE 4.14 Comparison of the infrared spectrum (from Gemini-South) of the electromagnetic counterpart to the binary neutron star merger GW170817 (black) to theoretical models of radioactively powered kilonova emission (red; see Ch 2.2). In the models, based on theoretical atomic structure calculations and radiation transfer, the broad absorption features are produced by a collection of highly Doppler-shifted transitions of neutron-rich heavy elements including neodymium and cerium. More detailed laboratory data on the atomic transitions of these and other neutron-rich elements are needed for a complete understanding of the spectra of neutron star merger counterparts and the heavy elements they produce. The region of strong absorption by Earth's atmosphere is indicated by the gray box. SOURCE: From R. Chornock et al 2017, "The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. IV. Detection of Near-infrared Signatures of r-process Nucleosynthesis with Gemini-South," *The Astrophysical Journal Letters*, 848 L19. © AAS. Reproduced with permission. doi:10.3847/2041-8213/aa905c.

Laboratory astrophysics was identified in the Astro2010 decadal survey report as “vital for optimizing the science return from current and planned facilities,” especially in the ALMA and JWST era.[23] Yet they found that “support and infrastructure for laboratory astrophysics are eroding both in the National Laboratories and in universities” and they recommended that “the funding through APRA that is aimed at mission-enabling laboratory astrophysics should be augmented at a level recommended by this scientific assessment . . . a notional budget increment of $20 million over the decade may be required.”[24]

Currently laboratory astrophysics is supported through grants from the NASA APRA and ADAP programs and NSF AST, as well as some support from DOE and national labs for laboratory astrophysics. However the number of awards approved across all of these programs is small and they have been declining since the early 2000s.[25] A search of NSF AST awards over 2015-2019 revealed 15 grants funded in laboratory astrophysics for a total of $6.2 million, and $12.4 million in funding from APRA over the same period. Despite the Astro2010 recommendation above for a $2 million per year increase in APRA funding for laboratory astrophysics, grant funding has remained essentially constant over the decade.[26] Given the growing need for laboratory data and the relatively small investment required relative to the costs of the facilities supported, enacting the Astro2010 recommendation is more important than ever.

It is important to add however that simply allocating more grant funding by itself will not be sufficient to address the entire problem. This research is most effective when the laboratory researchers have close ties to the astrophysical users of the experiments and data, but high start-up costs (typically $2 million or more) and the cross-disciplinary nature of the subject often leave university astronomy departments reluctant to hire new faculty in this area. Agency support for early-career faculty, similar for example to the NSF Faculty Development in Space Sciences (FDSS) program, could incentivize departments to invest in this field. Coordination and high-level prioritization of prime areas for future funding could also be effective. NASA for example already facilitates such an exercise through its Laboratory Astrophysics Workshop, but most of the resulting priorities are set by the researchers in the field. Broadening a similar exercise to include the user communities for the laboratory and computational data would be an important step towards ensuring that the precious funds are optimized to address the most pressing needs for interpreting current and future observations. Finally, for large flagship missions and MREFC-scale NSF facilities which rely heavily on laboratory data, including provision for these essential activities into the project budgets could be very cost effective and would naturally focus the laboratory work on the most urgent scientific needs for those facilities.

Conclusion: Laboratory astrophysics is essential to the interpretation of astrophysical data from facilities such as JWST, ALMA, and future facilities like the ELTs. Research in this area needs to be regarded as a high priority. The existing approaches are not sufficiently advancing the field.

Recommendation: NASA and the National Science Foundation should (1) convene a broad panel of experts to identify the needs for supporting laboratory data to interpret the results from the new generation of astronomical observatories, (2) identify the national resources that can be brought to bear to satisfy those needs, and (3) consider new approaches or programs for building the requisite databases. This panel should include experts in laboratory astrophysics as well as representative users of the data, who can best identify the highest-priority applications.

---

[23] NWNH, p. 32.

[24] NWNH, p. 220-221.

[25] See Section 4 of Nave et al. 2019; Atomic data for astrophysics: Needs and challenges. Bulletin of the AAS 51(7).

[26] 2018 NASA Laboratory Astrophysics Workshop: Scientific Organizing Committee Report; https://baas.aas.org/pub/2020i0202/release/1?readingCollection=1bea0260.

4.6 SUMMARY

"Research is formalized curiosity. It is poking and prying with a purpose. It is a seeking that he who wishes may know the cosmic secrets of the world and they that dwell therein. "
Zora Neale Hurston, Dust Tracks on a Road (1942)

This report lays out a roadmap for reaching the destinations for discovery in Chapter 2, A New Cosmic Perspective. Just as humans are fundamental to the success of research endeavors (as argued in Chapter 3), so too is the infrastructure for supporting that "poking and prodding with a purpose." These are foundational components to the astronomical research endeavor, without which no steady footing can be assured. Important components of this foundation are threatened, however, due to perennial underinvestment. Writing the present chapter brought to light multiple examples of previous decadal survey recommendations that remain unfulfilled. While the weakness of vital parts of the foundation has not prevented the extraordinary scientific advances of the past decade, as with any foundation, continued neglect and erosion of the foundation will continue to undermine the entire enterprise over the longer term. Realizing the opportunities that can be achieved with appropriate funding and focus is the best way to ensure that the science destinations are reached, so that new courses can be charted efficiently.

Ultimately understanding the connected cosmos through that "formalized curiosity," and reaching the ambitious decadal goals—unveiling the drivers of galaxy growth, new windows on the dynamic universe, and pathways to habitable worlds—requires more than big new machines. It requires people to translate observations into discoveries, theoretical studies to connect the observational clues, experiments in the laboratory and with the computer to interpret the data and the theory, and digital libraries of these precious data which meet the needs for the twenty-first century. Finally, support for big machines and big projects needs to be balanced with support for the individual researchers who are the wellsprings of scientific creativity and discovery. A few well-targeted, modest investments in the enabling research foundation will restore a healthy balance to the overall portfolio and maximize the scientific return.

5

Evaluating and Balancing the Operational Portfolio

Whereas Chapter 4 describes the research infrastructure, this chapter focuses on the suite of currently operating telescopes and missions that drives the scientific advances of today. Fully capitalizing on this suite of facilities requires managing and balancing the resources required to operate and maintain them, and upgrading capabilities where needed, in a way that returns maximum scientific benefit. This chapter draws from all of the program panel reports, and in particular from the Enabling Foundations panel report.

Although the largest component of national investment in astronomy goes to the development and construction of future major observatories and facilities (Chapter 7), today, the currently operating facilities on the ground and in space are the primary tools for collecting data that drives scientific discovery and progress in astronomy and astrophysics. Support for publicly shared facilities comprises the largest fraction of the NSF Division of Astronomical Sciences (AST) annual budget,[1] and a large share of the NASA Astrophysics Division budget that is not devoted to mission development. It is this funding that keeps observatories such as the Hubble Space Telescope (HST), Chandra, Fermi, the Transiting Exoplanet Survey Satellite (TESS), the Neil Gehrels Swift Observatory, the Nuclear Spectroscopic Telescope Array (NuSTAR), the Atacama Large Millimeter/submillimeter Array (ALMA) (the U.S. share), the Jansky Very Large Array (JVLA), and the Gemini Observatories among others running and delivering cutting-edge science. Collectively such facilities have been extraordinarily productive (see Chapter 2) and cost effective in seeding a steady stream of major scientific discoveries using panchromatic capabilities that are central to their advancing broad decadal scientific priorities.

The vitality of these facilities is routinely assessed, with NASA and NSF engaging in periodic reviews of their portfolios of operating missions and facilities. The importance of evaluating the operational mission/facility portfolios on a regular basis was underscored by the 2000 and 2010 astronomy and astrophysics decadal surveys, and for NASA in a 2016 National Academies study of NASA mission extensions and the senior review process.[2] All of these studies emphasized the importance of such reviews to optimize the scientific return on these facilities investments.

As with the assessment of the research foundation activities in Chapter 4, the main interest of this survey is not with the details of these stewardship processes, but rather in assessing at a high-level which aspects of the processes are functioning effectively and which are less healthy, due to factors such as growing programmatic imbalances, unforeseen events, or from rapid changes in the overall scientific landscape or priorities since the last decadal survey. As part of this assessment, the committee considered not only the processes for prioritizing individual missions and facilities within the agency portfolios, but also the overall balances of investment between facilities and the many other types of investments that are also needed to advance science.

Before addressing NSF and NASA portfolios individually, it is worth highlighting general areas in which this committee believes the current management of operating facilities has been particularly

---

[1] NSF FY2021 Budget Request, https://www.nsf.gov/about/budget/fy2021/pdf/fy2021budget.pdf.

[2] National Academies of Sciences, Engineering, and Medicine, 2016, *Extending Science: NASA's Space Science Mission Extensions and the Senior Review Process*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/23624.

successful. The NASA Astrophysics Senior Review of Operating Missions has proven to be especially effective in setting funding priorities for operating missions (post-prime mission), and for establishing criteria and a decision process for terminating missions. Its 3-year cadence ensures that all projects regularly document their scientific productivity, user demand, data products, operational plans, and budget allocations on a regular basis. In contrast, the NSF's Senior and Portfolio Reviews are conducted less frequently, and without a predictable cadence (last conducted in 2006 and 2011-12).

This committee however did identify areas of concern, where imbalances or inconsistencies across the agency portfolios now pose threats to the overall science return, vitality, and sustainability of the astronomy and astrophysics programs. For NSF, the chief concern is insufficient funding to support operations of high-impact scientific facilities that are at the core of astronomy's current and future ground-based research enterprise. The result, as detailed in Chapter 4, is a trend toward a declining fraction of the AST budget for other purposes, most notably the Astronomy and Astrophysics Grants program (AAG). The problem is poised to become much worse with the imminent commissioning of the Rubin and the Daniel K. Inouye Solar Telescope (DKIST) observatories and is an existential concern when contemplating the exciting set of proposed Major Research Equipment and Facility Construction (MREFC) projects (U.S. Extremely Large Telescope [ELT] program, the Next Generation Very Large Array (ngVLA), the Stage-4 ground-based cosmic microwave background experiment [CMB-S4]), and the sustaining instrumentation program recommended by this survey (see Chapter 7). For NASA, a serious concern is the exclusion of a major facility (the Stratospheric Observatory for Infrared Astronomy [SOFIA]) from the Senior Review process. This committee addresses these and a few other, less serious concerns separately for each agency.

5.1 NSF OPERATIONAL FACILITIES

This section provides an analysis of and recommendations related to NSF's current model for the operation of major facilities, along with assessments and recommendations on maintaining its current portfolio, especially in the OIR.

5.1.1 NSF Funding for Major Research Facilities

In contrast to budgeting for NASA space missions, that include end-to-end funding for construction, launch, and operations through the prime mission phase, the NSF budgets instead separate the funding streams for facility construction and operations. Under current NSF regulations, the construction of projects that cost more than $70 million may be funded by the agency-wide MREFC program. Proposals to the program are based on design and development efforts funded by a division and/or directorate, and, if the proposals are accepted, the MREFC program takes over and provides processes for planning, oversight, and review throughout the construction process. This structure is ideal for astronomy, with its reliance on transformative, widely-shared facilities.

However, while the MREFC process has supported building revolutionary facilities like ALMA and Rubin, the program does not provide support beyond construction, leaving the operations and maintenance (O&M) costs of these facilities as the responsibility of the sponsoring directorate, but without a commensurate, sufficient, increase in the directorate's funding line to account for operations costs. With many NSF facilities having lifetimes of 50+ years, and annual operations costs typically amounting to 4-7 percent of the original construction cost,[3] the total lifetime cost of O&M can easily rival or exceed the original cost of construction. Moreover, the O&M costs are typically not carried by the directorate as a whole, and instead are passed down to an individual division. For astronomy facilities,

[3] B. Goodrich, C. Dumas, M, Dickinson, R. Bernstein, P. McCarthy, 2019, Observatory operating costs and their relation to capital costs, APC white paper submitted to the Astro2020 decadal survey.

these costs are almost always borne by AST, with occasional contributions from Physics or Polar Programs. The division must then carry these costs for the remainder of the productive scientific lifetime of the observatory.

Research communities from across NSF's divisions have expressed mounting concern about the impact of the "O&M mortgage" on the overall health of their fields. In response, the U.S. Senate Committee on Appropriations issued guidance to the National Science Board (NSB) in its FY 2017 appropriations bill for the Departments of Commerce, Justice, Science and Related Agencies (S.2837):

> Operations and Maintenance Costs—The Committee is concerned that operations and maintenance costs for NSF-funded research facilities require an increasingly large percentage of the funding for Research and Related Activities, especially in a budget environment where overall domestic spending is restrained and annual operations and maintenance costs increase faster than overall NSF spending. The Board is directed to consider whether this issue merits a change in NSF's funding principles or budgetary formulation processes, including considering the research infrastructure funding approaches within other Federal agencies, and whether a separate operations account is merited.

The NSB responded to this charge with a 2018 report (NSB-2018-17) entitled *Study of Operations and Maintenance Costs for NSF Facilities*. This report found that in nearly all facilities-heavy divisions and directorates, including AST, O&M spending has increased faster than division and directorate budgets. Figure 5.1 shows the fraction of the division budgets represented by O&M costs for several different divisions. The report notes that in divisions other than AST (Physics, Materials Research, Earth Sciences, Geosciences, Ocean Sciences, Earth Sciences), this fraction has leveled off at below or around 30 percent. AST stands out among all other divisions as having facility O&M costs that are projected to continue to rise.

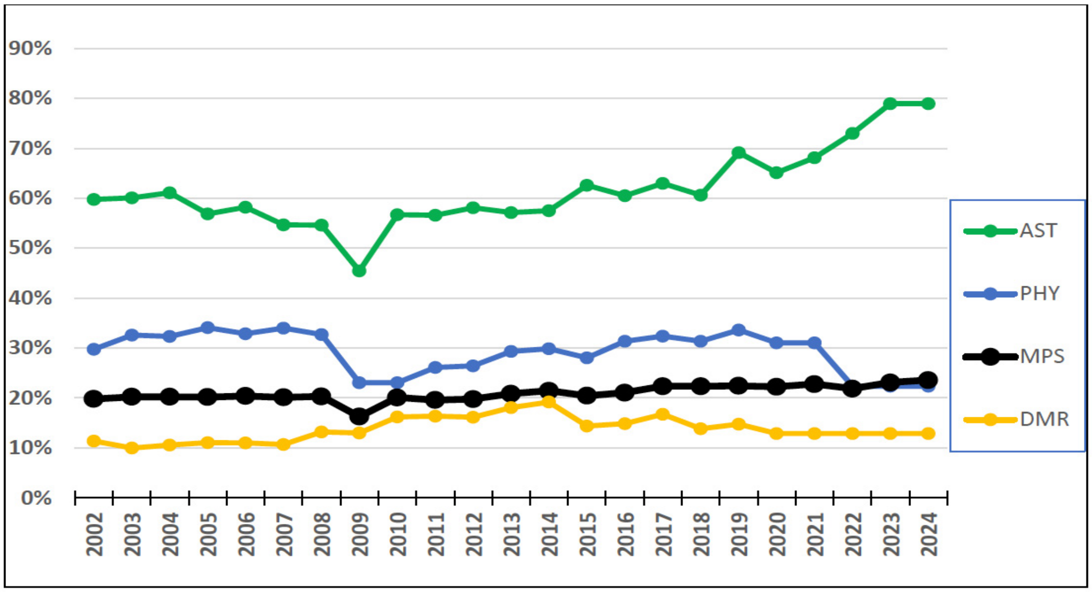


Figure 2, Percentage of Selected MPS Division Budgets to Facilities (O&M) and Overall MPS Share. AST: Division of Astronomical Sciences, PHY: Division of Physics, MPS: Directorate of Mathematical and Physical Sciences, DMR: Division of Materials Research

FIGURE 5.1 Percentage of Selected MPS Division Budgets to Facilities (O&M) and Overall MPS Share. Budget numbers through 2017 are actuals. Budget numbers from 2018 on are projections. The projected fraction for AST has been roughly consistent with this analysis up to the present date. NOTE: AST: Division of Astronomical Sciences, PHY: Division of Physics, MPS: Directorate of Mathematical and Physical Sciences, DMR: Division of Materials Research. SOURCE: NSB report NSB-2018-17. Courtesy of the National Science Board.

The two divisions with the highest O&M budget fractions are AST and Oceanography (OCE), and this is no coincidence. Both fields rely heavily on shared national research facilities; in the case of AST mainly radio, submillimeter, and optical-infrared ground-based observatories, and for OCE large oceanographic vessels. This research infrastructure model contrasts sharply with other divisions such as physics, chemistry, materials science, etc., where the bulk of research facilities reside within individual university and institutionally-based laboratories.

The main components of the current AST O&M portfolio are the National Radio Astronomy Observatory (NRAO) (including the U.S. share of ALMA and the JVLA), the National Optical-Infrared Astronomy Research Laboratory (NOIRLab) (including the two Gemini telescopes, Kitt Peak National Observatory [KPNO], Cerro Tololo Inter-American Observatory [CTIO], and Datalab), the National Solar Observatory (NSO), Green Bank Observatory (GBO), Arecibo (See Section 5.1.5), and some elements of the AST Mid-Scale Innovations Program (MSIP, discussed separately in Section 6.3.2). These facilities are all highly valued by the community, but their aggregate O&M needs have imposed a severe squeeze on the rest of the AST budget. Programs particularly impacted are the AAG individual grants line (see Section 4.2.1), support for technology through ATI (see Section 6.1.2), instrumentation, and graduate education. Other elements of the program affected by the squeeze include support for technology, instrumentation and graduate student education.

Soon, two new major MREFC facilities recommended by the Astro2000 and Astro2010 decadal surveys, DKIST and the Vera Rubin Observatory, respectively, will place even greater strain on the AST budget. As stated in the 2018 NSB study:

> The Division of Astronomy (AST): The Division of Astronomical Sciences situation is highlighted on p. 21-22 of the NSB report, which notes: "[W]ith limited budget growth, the almost $100 million in steady-state O&M needed when three state-of-the-art facilities that were, or will be, completed between 2012 and 2023 is challenging the division's ability to manage its portfolio of existing and future facilities without severely affecting its investigator research program."

The NSF AST has long been aware of the problem, and has attempted to adjust in response to rising facilities costs. AST undertook a portfolio review in 2011-2012, that was charged with examining how the program recommended by Astro2010 could be realized within a more limited budget profile than anticipated. The portfolio review recommended a course of divestment from a number of facilities. Some of these divestment recommendations were adopted, and provided a total cost savings of about $15 million per year. More extensive divestment of legacy, but scientifically productive, facilities, could generate an additional savings at this scale but are insufficient to compensate for the needs of upcoming facilities and at the same time fund individual investigator grants at a healthy level. As stated in the midterm assessment of the Astro2010 decadal survey, "divestment alone will not resolve the budget stresses imposed by rising facilities costs."[4] The midterm assessment report appealed to NSF and NSB to "consider actions that would preserve the ability of the astronomical community to fully exploit the foundation's capital investments in ALMA, DKIST, LSST, and other facilities. Without such action, the community will be unable to do so because at current budget levels the anticipated facilities operations costs are not consistent with the program balance that ensures scientific productivity."[5]

As this committee assessed the ambitious set of proposed MREFC projects for the coming decade, it became clear that to implement any of them, and at the same time support continuation of the world-leading observatories such as Rubin, DKIST, ALMA, the JVLA, and Gemini, requires a fundamental change in the budgets available for AST O&M. The major projects presented to this survey carry capital costs to the MREFC line ranging from hundreds of millions to approximately $2.5 billion

[4] National Academies of Sciences, Engineering, and Medicine, 2016, *New Worlds, New Horizons: A Midterm Assessment*, Washington, DC: The National Academies Press, https://doi.org/10.17226/23560.

[5] Ibid.

dollars, with corresponding operating costs, ranging from $20 million to $100 million per year (for any one facility, as estimated by the projects themselves). When compared to the approximately $80 million currently available in the AST budget outside of the O&M line, the current challenge has the real potential to escalate the existing funding problem in the AAG program (Chapter 3) into a full crisis in the near future, when DKIST and Rubin are fully operational. Although some of the pressure will be relieved by planned phased retirements of current facilities (for example retirement of the JVLA when the ngVLA comes into operation), it will not by itself relieve the existential threat to the sustainability of the NSF astronomy budget.

In short, the structural difficulty of funding on-going operations of long-lived, scientifically productive astronomy facilities places a profound challenge in front of the roadmap laid out by this decadal survey. The same transformative projects that could readily attract MREFC funding would simultaneously make it impossible to actually carry out the science, because of the inevitable underfunding of research grants to use the new facilities, and of the theoretical studies and computational tools needed to harness and interpret the data. Advancing astronomy and astrophysics takes the cutting-edge facilities, the means to analyze and interpret the data, and also the theory and open-ended ideas that will make the leaps from data to discoveries.

The only ways out of this dilemma are through augmentations to AST's overall budget, or through changes to the current NSF model for funding of construction and operations of large facilities. It is imperative that the agency work with the AST division and MPS directorate to develop a *sustainable* budget and/or model for construction and operation of new facilities, one that allows our community to maintain an appropriate balance of investments in all of the other critical elements of the enabling foundation for research that have been outlined in Chapter 4.

The 2018 NSB report makes several recommendations aimed at achieving this objective. One key recommendation is stronger agency level oversight and involvement in strategic planning for major facilities. A longer timescale for budgetary planning (currently facility budgets contain 5-year projections) is also suggested. Notably, the report discusses a vision for a more flexible implementation of the MREFC account, under which partial funding for O&M for a new facility could be allocated from MREFC for a limited period (5-10 years). O&M costs could then be gradually absorbed into a division or directorate budget. This welcome adjustment could help to solve the problem of operations costs for shorter-lived facilities, it would only temporarily alleviate but not eliminate, the funding pressures from Astronomy's existing capital investments. Other solutions, such as creating an operations budget line at the MPS or AST levels sized to accommodate O&M for current facilities, and the planned profile of which would anticipate future needs, would also address this issue.

Conclusion: The current pressure imposed by operations costs of large NSF facilities on the grants and other NSF programs will escalate to unsustainable levels by mid-decade unless changes are made to the way that large facilities are supported.

Recommendation: The National Science Foundation (NSF) should develop a sustainable plan for supporting the operations and maintenance costs of its astronomical facilities, while preserving an appropriate balance with funding essential scientific foundations and the remainder of the NSF Division of Astronomical Sciences portfolio. The addition of new MREFC facilities should be contingent on implementation of this plan.

5.1.2 Managing the NSF Facilities Portfolio

As highlighted earlier, periodic reviews by NASA and NSF of their portfolios have proven to be effective mechanisms for maximizing science return and prioritizing budgets. The NASA Senior Review, which is undertaken every 3 years, has proven to be an extremely effective way to maintain high scientific productivity while managing costs. In response to a recommendation in the Astro2010 decadal survey,

NSF organized a portfolio review of its operating facilities in 2011-2012, but none since. Reviews on a regular cadence allow for a periodic evaluation of the productivity and science return from facilities, and can help identify where efficiencies can be realized, or where funding augmentations might be required to capitalize on new scientific opportunities. Such reviews can also identify those observatories from which NSF might divest and subsequently decommission. These goals largely mirror those carried out in the 2011-2012 portfolio review, and this is a practice worth implementing on a regular basis to ensure NSF facilities regularly assess, document, and justify their performance and service to science and education.

Similarly to the NASA Senior Review process, NSF facility reviews would focus on scientific promise, productivity, and budgetary efficiency. The committee appreciates that some aspects of facility reviews have taken place as parts of the review of operating agreements for observatories, but such reviews are not an appropriate substitute for a review which considers the entire portfolio simultaneously on a holistic basis. The cadence of NSF reviews need be sufficiently frequent to allow facilities to respond to changes in the scientific and budgetary landscape in advance of the decadal and mid-decadal processes (i.e., at least twice a decade), but not so frequent that the facilities do not have time to implement and evaluate changes made in response to a previous review (i.e. not as frequent as every 2 years).

Recommendation: The National Science Foundation Division of Astronomical Sciences should establish a regular cadence of reviews of its operational portfolio, at a frequency that is sufficient to respond to changes in scientific and strategic priorities in the field. An appropriate target is at least two reviews per decade.

5.1.3 Investment in Mid- and Small-Scale Ground Facilities

Over the last decade, the system of ground-based telescopes in the United States has evolved significantly, in both the radio and OIR parts of the electromagnetic spectrum. Radio astronomy has had federal funding support the construction and operation of ALMA, an upgrade to the JVLA, and a collection of smaller specialized telescopes, collaborations, and instrumentation, albeit at the expense of the closing of the Combined Array for Research in Millimeter-wave Astronomy (CARMA) and the elimination of the University Radio Observatories program at NSF. In the OIR, key developments include the construction of the national flagship telescopes DKIST and Rubin, the innovative repurposing of several smaller ground-based telescopes (e.g., the Dark Energy Camera [DECam], the Dark Energy Spectroscopic Instrument [DESI], and the Zwicky Transient Facility [ZTF]). In spite of the investments in cutting-edge smaller facilities and experiments, capable radio and OIR observatories are currently aging and with insufficient investment are quickly becoming less competitive on the world stage, even among some of our largest facilities.

The operating ground OIR and radio facilities play vital roles, both in supporting observations with larger flagship facilities and generating major discoveries on their own. They are fundamental to preserving a balanced portfolio of capabilities and investments in U.S. astronomy and astrophysics. Flexible and accessible, ground-based telescopes can respond to a rapidly changing scientific environment through technology upgrades in cameras and detectors.

Critical to the future effectiveness of ground-based facilities are updates to instrumentation. Technology advances significantly over the decades-long lifetimes of these facilities, as do the needs of the scientific community. A robust investment in instrumentation upgrades can enable an observatory to maintain its competitiveness for far less than the cost of building a new observatory. The development of adaptive optics (AO) in the 1990s serves as an excellent case in point, and led to ground-breaking advances such as the direct imaging of exoplanets, and the precise definition of the orbits of stars that determined the gravitational force fields near the black hole at the center of the Milky Way, work recognized with the 2020 Nobel Prize in Physics. The upgrade of the receivers at the JVLA led to a factor of 10 increase in sensitivity for continuum observations at the higher frequencies, and nearly complete access to the 1-50 GHz frequency range.

NSF programs for supporting new and upgraded instrumentation are available at scales ranging from the MSIP and Mid-scale Research Infrastructure (MSRI) mid-scale opportunities, down to the Advanced Technologies and Instrumentation (ATI) program which supports smaller projects (see Chapter 6). Over the past decade these programs have supported a wide range of meritorious projects, including for example the Event Horizon Telescope (EHT), ZTF, and a number of cosmic microwave background (CMB) projects.

New and upgraded instrumentation on telescopes of all scales has also been supported significantly by private foundations and philanthropy. Investments by foundations such as Moore, Heising Simons, Sloan, Keck, Packard and others have been crucial, often in partnership with federal funding to developing new, ambitious instruments as well as repurposing smaller telescopes for targeted objectives. A few of many examples of these public-private instrumentation partnerships include the Las Cumbres Observatory, the Zwicky Transient Facility, the PolarBear/Simons Array, BICEP-Keck Array, the Keck Cosmic Reionization Mapper, and the Keck Planet Finder.

When assessing the current balance across these programs the committee identified a notable gap in the support for instrumentation on OIR telescopes. Previously some of this gap was filled by the Telescope System Instrumentation Program (TSIP), in which NSF provided money for development on private facilities in exchange for public observing time. During 2002-2011, NSF invested $33 million in 19 instrumentation projects for five observatories (Keck, the MMT Observatory, Magellan, the Wisconsin-Indiana-Yale-NOAO [WIYN] Observatory, and the Large Binocular Telescope [LBT]), and providing in return 453 nights of public access observing time on seven telescopes (the Keck and Magellan Observatories each have two telescopes), which was allocated through the National Optical Astronomy Observatory (NOAO) time allocation process.

However, TSIP was phased out early in the last decade and replaced by the MSIP program. Notably, unlike TSIP, MSIP does not require a return of public access in exchange for supporting instrumentation on private telescopes. A number of MSIP projects did provide public access time, for example 40 Keck telescope nights over 4 years resulting from funding for the Keck Planet Finder, 60 public nights per year on the Center for High Resolution Astronomy (CHARA) array, 2840 hours per year on the Las Cumbres Observatory global telescope network, and public-access targets of opportunity observations on ZTF,[6] the latter two mainly for time-domain applications. Applications for time and the time allocation processes are coordinated by NOIRLab through a common once-per-semester process. However, overall, since the replacement of TSIP with MSIP, the number of publicly-available nights on leading facilities such as Keck and on capable 2-4 m-class telescopes have decreased significantly.

Strategic use of ground-based telescopes has also proven to be advantages for supporting the scientific goals of NASA space missions, and, as a result, NASA has now joined NSF as a major investor in ground-based OIR observatories and instrumentation. As a partner in the Keck Observatory NASA allocates 1/6 of the time on the two 10 m telescopes (100 nights per year total) for public access.[7] The NASA-NSF Exoplanet Exploration Program (NN-EXPLORE) has funded construction of exoplanet-focused instruments and made available telescope time on the WIYN (adding a powerful new Doppler spectrograph), the Anglo-Australian Observatory (AAO), the Small and Moderate Aperture Research Telescope System (SMARTS), and Miniature Exoplanet Radial Velocity Array-Australis (MINERVA-Australis) observatories. NASA also operates the 3 m Infrared Telescope Facility (IRTF) for astronomical and planetary-science observations. NASA funded construction of the Large Binocular Telescope Interferometer (LBTI), and in this case 40 nights of public time were awarded for a single key project, *The Hunt for Observable Signatures of Terrestrial Systems (HOSTS) survey*, comprised of a nationally-competed team of investigators.

DOE is another significant contributor to U.S. ground OIR instrumentation, albeit focused on areas aligned with Office of Science objectives. In addition to providing the focal plane camera for Rubin, DOE has funded two extremely powerful optical survey instruments, DECam and DESI, using existing

---

[6] http://ast.noao.edu/observing/call-for-proposals-2021b
[7] https://nexsci.caltech.edu/

NOAO 4 m telescopes. DOE's investments are motivated by the central importance of large-scale surveys for cosmology, and the instrument designs are in large part optimized for this purpose. However, the resulting data have wide astronomical utility.

These contributions from NSF, private foundations and philanthropy, NASA, and the DOE have helped to maintain the vitality of the instrumentation on ground-based OIR telescopes, but fall short of what is needed to maintain the competitiveness of the facilities. As one useful benchmark, ESO currently invests ~$10 million per year in instrumentation across its OIR telescopes, with another $15 million to $20 million invested by the partner organizations. Two major instruments on the Japan-led Subaru 8.4 m telescope, the Hyper Suprime-Cam and the future Prime Focus Spectrograph, have associated costs of $50 million and $86 million, respectively. Funding even shares of such instruments will require larger allocations than historically have been awarded through the TSIP or MSIP programs. The second challenge is the loss of public access time over the past decade, arising both from the discontinuation of the TSIP program and the effective withdrawal of the CTIO 4m Blanco and KPNO 4m Mayall telescopes from general public use during the course of the DES and DESI (and their associated) multi-year surveys. The point is not to criticize whatsoever the decisions to undertake these important surveys, but rather to emphasize the ever-dwindling general public access to 3.5—10m class OIR telescopes in recent years.

Although NSF's MSRI program is an excellent new opportunity for funding larger ground-based astronomy instrumentation projects, it will be difficult to sustain the required level and cadence of new instrumentation through the program as currently planned. Given the significant investments, both public and private, in OIR telescopes, these facilities will remain a critical part of the U.S. ground-based astronomy program for decades, and over time will need instrumentation upgrades to remain at the cutting-edge. To achieve this, in Chapter 7 we recommend periodic, strategic calls through AST and NSF mid-scale program lines specifically to support upgrades of instrumentation on OIR facilities (private and public), both to maintain the scientific capabilities of those facilities and as a mechanism to expand community access to them. This includes a provision that such awards carry a requirement for allocating public observing time on the facilities, along the lines of the previously successful TSIP model, and for public release of data from the relevant instruments (after a suitable proprietary period). These mechanisms may not be suitable for some survey projects of experiments, but the broad objective or assuring community benefit from the federal investments should be met to whatever extent possible.

> **Conclusion:** U.S. competitiveness internationally in ground-based OIR astronomy requires a stable funding mechanism for instrumentation development on existing ground-based telescopes that includes public access for the community.

5.1.4. Opportunities for Maximizing Public Investments

The U.S. system of ground-based astronomical observatories includes both federally funded facilities and facilities constructed and operated by academic and private institutions. In the radio, millimeter, and submillimeter, telescopes are typically federally funded, primarily by NSF AST. In the OIR, however, many of the most advanced and powerful ground-based observatories are funded and operated by consortia of academic institutions and private foundations. This funding model favors access by astronomers who are affiliated with the participating organizations which invest in the construction and support of those facilities. Another manifestation of this de-centralized system is a lack of coordination across the many observatories, whether it be in terms of setting common priorities for new instruments, arranging for exchanges of observing time to limit duplication of instrumentation, or in representing their common interests with the agencies.

The concept of a more coordinated system of ground-based observatories is hardly a new idea. On the recommendation of the Astro2000 decadal survey, NSF established the concept of the "OIR system," which was intended to balance and optimize coordination of the national and private observatories. Ten years later the NWNH report concluded that "Optimizing the long-term science return

from the whole of the U.S. optical and infrared system requires a readjusting of the balance of the NSF-Astronomy program of support in three areas: (1) publicly operated national observatories; (2) private-public partnerships—such as support for instrumentation and upgrades of privately operated observatories; and (3) investment in future facilities." Following both reports, NSF organized strategic planning committees to develop decadal roadmaps for optimizing the OIR system, and these led to significant advances, not the least the coordination of public access to privately-run facilities as described earlier.

There is value in expanding this model beyond an ad hoc committee which issues a report once per decade, towards a standing committee which would facilitate dialog between the diverse set of OIR stakeholder institutions, and serve in the coordinating role envisaged above. The newly formed NOIRLab may be an entity that could convene (but not direct) such an activity.

> **Conclusion:** As the cost of new instrumentation on ground-based OIR telescopes continues to increase, improved coordination and collaboration among facilities run and/or supported by private institutions, universities, national laboratories, and private foundations could facilitate the development of a coherent national strategic plan for OIR astronomy.

5.1.5 The Arecibo Observatory

In December 2020, the storied Arecibo radio telescope in Barrio Esperanza, Arecibo, Puerto Rico collapsed, ending a remarkably productive 55 plus year service to the astronomical, planetary, ionospheric, defense, and space-science communities. Innovative in design, its large, fixed dish with a suspended secondary enabled the discovery of over 500 pulsars, including the Hulse-Taylor binary pulsar that led to the discovery of gravitational radiation and the 1993 Nobel Prize in physics. Other notable contributions include: the first discovery of an exoplanet that was found orbiting the pulsar 1257+12; the search for extra-terrestrial intelligence (SETI); mapping of hydrogen gas emission over thousands of square degrees; understanding the composition of the ionosphere; the characterization of the properties and orbits of a number of potentially hazardous asteroids; radar mapping the surfaces of Mars, Venus, and Mercury (including its ice); and most recently probing Fast Radio Bursts (FRBs) and intermittent pulsars to refine our understanding of the underlying physical mechanisms. At the same time, through its Ángel Ramos Foundation Visitor Center, it brought the wonders of the radio sky to many hundreds of thousands of students and non-specialists. Arecibo was also a focal point for STEM-related education in Puerto Rico, inspiring a young, diverse generation to pursue a career in science.

Arecibo opened a new, rich, view of the cosmos that over the years has helped to spark new ideas and more ambitious efforts. The results of its HI surveys of the Milky Way Galaxy and the census of the local HI content of nearby galaxies has motivated the larger efforts of the Square Kilometer Array (SKA) and its precursors. Furthermore, the Five-hundred-meter Aperture Spherical Telescope (FAST) in Guizhou, China builds on Arecibo's 305 m diameter. Since first light in September 2016, FAST, which covers from 70 MHz to 3 GHz as compared to Arecibo's 50 MHz to 10 GHz, has discovered over 200 new pulsars as of May 2021.[8] For specific types of astronomical searches in much the same frequency range, other instrumentation approaches are better. For example, the Canadian Hydrogen Intensity Mapping Experiment (CHIME, 400-800 MHz) is just 80 m^2, yet has greater mapping speed than FAST. The Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX) will complement CHIME in South Africa. The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD; 0.3-1.5 GHz, 512, 6 m dishes) and the proposed DSA-2000 (0.7-2 GHz, 2000, 5 m dishes) are expected to discover, for example, many thousands of new pulsars and FRBs. This new generation of telescope blends modern high-speed digitization and correlation of multiple resolution elements with optimized hardware to simultaneously monitor large swaths of sky, acting like radio cameras. To complement these, in Chapter

[8] Han et al. http://www.raa-journal.org/raa/index.php/raa/issue/view/231

7, we recommend commencing with the development of the ngVLA, an international effort to construct an array of more than two hundred 18 m telescopes operating from 1.2 to 116 GHz. While none of these new telescopes individually will replace all the capabilities of Arecibo, the combination of the radio cameras and ngVLA will be much more powerful for broadly advancing astrophysics in this frequency range.

The survey steering committee assessed the impact of Arecibo's loss on the key science questions and program elements forwarded by the Astro2020 panels, while noting that these topics are largely outside the planetary and solar system science fields where Arecibo has had tremendous impact. When restricted to astrophysics alone, the most significant impact is the loss of the Arecibo's contribution to discovering and timing pulsars that are elements of the Pulsar Timing Array (PTA). This set of pulsars is used to search for new sources of gravitational radiation. The PTA reveals gravitational waves with ~year periods through slight alterations in the arrival times of the emission from a catalog of millisecond pulsars. To date, over a quarter of these pulsars have been discovered with Arecibo and timed with telescopes including Arecibo, the Green Bank Telescope (GBT) and the JVLA.

To reach the scientific goals of Astro2020, much of what was lost with Arecibo can be replaced, in the near term, by focusing more resources on timing with the JVLA and GBT, and by increased collaboration with the international community undertaking pulsar searches. As pointed out by the RMS panel, the uniqueness of large single-dish telescopes like Arecibo, or very closely packed arrays, is their ability to search for new sources to improve the sensitivity of the PTA to gravitational waves. As noted above, the FAST telescope is already filling this role, and expanded international collaboration could ensure continued detection of relevant new sources. For pulsar timing, additional observing time will be needed on the GBT, and also the JVLA, to provide the necessary phase-connected timing solutions. On few-year timescales, radio cameras will add search capabilities. This survey recommends commencing development of the ngVLA this decade and also recommends adding radio instruments as a strategic call at the mid-scale (Chapter 7). Therefore, in the longer-term, new facilities will advance a broad range of Astro2020 science goals, including the detection and study of FRBs, and pulsar timing.

Another important goal of Astro2020 is to enhance community engagement with astronomy. The Ángel Ramos Foundation Visitor Center has been a model for this. In Chapter 3 we address the importance of local community involvement in realizing the goals of Astro2020. In addition to this education and public outreach component, the observatory promotes demographic diversity in STEM through its impact on post-secondary education. These activities are important and worth continuing. Looking to the future, the reference design for the ngVLA calls for at least one of its antennas to be placed in Puerto Rico, an example of one path for Puerto Rican communities to become part of a connected network of telescopes that span from Hawaii to the Virgin Islands, and that will at the same time be on the forefront of astronomical research and discovery.

Finding: There are future opportunities for continued utilization of the Arecibo site for radio astronomy, both through the ngVLA and mid-scale projects.

In summary, Astro2020 took a broad view of all the capabilities needed to ensure a strong future in radio astronomy. These recommendations are synergistic with the entire multi-messenger, multi-wavelength astronomy program. Even in the absence of Arecibo, and in light of multiple new ideas for its replacement (e.g, Roshi et al.),[9] our priorities for radio astronomy are: the support of existing facilities; the phased build-up toward the ngVLA; and competed small and mid-scale projects, all undertaken in an international context. Both because of its location and its communities, Puerto Rico has an important role to play in the future of radio astronomy, and it remains a good site for investing in mid-scale radio

[9] D. A. Roshi , N. Aponte, E. Araya, H. Arce, L. A. Baker, W. Baan, T. M. Becker, et al., 2021, "The Future of the Arecibo Observatory: The Next Generation Arecibo Telescope" white paper last updated on February 1, 2021, arXiv:2103.01367.

projects. We encourage NRAO/Associated Universities Inc. (AUI) and other entities to take advantage of this opportunity along the lines outlined in this report.

> **Conclusion:** Much of the science relevant to the Astro2020 goals lost with Arecibo can be recovered through additional investment in existing facilities, and through international partnerships, while the new facilities recommended by this survey are realized.

5.2 NASA OPERATIONAL FACILITIES

NASA has a well-defined and effective process, the Astrophysics Senior Review of Operating Missions, proven to be effective in setting funding priorities and for establishing criteria and a decision process for terminating missions. A 2016 NAS Report, "*Extending Science: NASA's Space Science Mission Extensions and the Senior Review Process*" found that across all of NASA's science programs, extended missions are an important part of both achieving decadal science objectives, and determining priorities or approaches for future exploration. This report also finds that senior review is the best mechanism for advising NASA about budgetary levels, or advising when a mission should be terminated because its scientific return is not commensurate with the requisite investment. As such, decadal surveys do not typically weigh in on individual operating missions. However, SOFIA was not considered by the last senior review panel, and the value of continuing operations of SOFIA beyond 2023 is of concern with respect to the other priorities of this report.

5.2.1 SOFIA

In the budget presented to Astro2020, NASA did not include plans to continue operating the Stratospheric Observatory for Infrared Astronomy (SOFIA) beyond 2023. NASA conducted two separate review exercises: the SOFIA Operations and Maintenance Efficiency Review (SOMER) and the SOFIA Five Year Flagship Mission Review (FMR). The decadal survey, as part of its overview of the current state of astronomy and astrophysics science and technology research, considered the outcome of these reviews and made its own evaluation of the relative scientific value of continuing SOFIA relative to Astro2020 science questions, and relative to other decadal survey priorities.

SOFIA, prioritized by the 1990 and 2000 decadal surveys, observes with an air-temperature 2.5-meter telescope mounted in a highly modified Boeing 747. It typically flies at altitudes over 11.3 km, which is above 99 percent of Earth's precipitable atmospheric water vapor, allowing for access to infrared wavelengths not possible from the ground. SOFIA's instrumentation has therefore focused on mid to far infrared, via both spectroscopy and imaging. The SOFIA project is joint between NASA and Deutsches Zentrum für Luft- und Raumfahrt e.V (DLR, the German Space Agency), with NASA providing 80 percent of operations costs and DLR 20 percent. The SOFIA program was started under contract with the Universities Space Research Association (USRA) in 1996, saw first light in May 2010 and achieved full operational capabilities in May 2014. SOFIA performs mostly northern hemisphere flights, as it is based in Palmdale, CA, and spends a smaller fraction of the year in the southern hemisphere, where it takes off from Christchurch, New Zealand.

The survey committee has significant concerns about SOFIA, given its high cost and modest scientific productivity. The NASA portion of SOFIA's operating budget is $86 million a year, of which $4 million goes to Guest Observers for data reduction and analysis. This yearly budget is in a range comparable to NASA's flagship space telescopes Hubble and Chandra ($98 million and $62 million in FY2019, respectively). The total life cycle cost for SOFIA to date is ~$1.5 billion. For this investment, the science productivity to date is very low: 178 total papers after 6 years (from May 2014 to May 2020). The science impact is also low; these papers have a relatively low citation rate: for the same time period, only 1242 citations. As a comparison, in the first 6 years after the launch of each of Hubble and Chandra,

with similar yearly budgets, the community produced more than 900 and 1800 total papers, respectively. Similarly, ESA's Herschel mission, was, like SOFIA, a flagship scale mid-to-far-infrared facility, which saw nearly 900 peer-reviewed papers in the 6 years following launch (170 papers in the first year alone), with more than 20,000 citations during the same period. Comparing to a more recent mission that has been operating for a shorter time than SOFIA, NASA's TESS –a Medium-class Explorers (MIDEX) mission—had in its first 2 years of operations (launch date April 2018) 281 papers with 2322 citations. In addition, SOFIA's clearly unique capabilities across these important wavelength ranges have not translated into high utilization of the observatory by the astronomical community. For instance, only 9 of the 35 SOFIA-related Ph.D.'s are from U.S.-based students, as of Fall 2019, and the single largest producer of SOFIA's scientific publications to date is Germany's Max Planck Institute for Astronomy. Furthermore, some of the originally motivating unique capabilities in 1990 and 2000 have since been superseded by the results from the Herschel Space Observatory (2009-2013). Relative to its cost, SOFIA has not been scientifically productive or impactful over its duration.

To assess SOFIA's potential impact going forward, we determined the role SOFIA could play in Astro2020 science priorities. We find that SOFIA directly addresses three of the thirty priority science questions (Question 4 of the Panel on Galaxies, Question 2 of the Panel on the Interstellar Medium and Star and Planet Formation [ISM], and Question 4 of the Panel on the Stars, the Sun, and Stellar Populations), and indirectly contributes to one more (Question 4 of ISM). There is therefore minimal overlap of the Astro2020 Panels' science priorities with SOFIA capabilities.

Some of SOFIA's challenges are inherently structural. The operation of the observatory is complex, and a large staff is required both to maintain the observatory and to perform observing runs. There is significant down time in each year for necessary airplane maintenance. With a typical ~1000 flight hours per year, and a relatively modest 60 percent of programs being completed, and 60 percent of these turning into peer-reviewed publications, only a few percent of total yearly calendar hours are turned into peer-reviewed science, an order of magnitude less than other astronomical observatories.

In 2018-2019 NASA charged two review committees to assess the state of SOFIA. These were the SOFIA Operations and Maintenance Efficiency Review (SOMER) and the SOFIA Five Year Flagship Mission Review (FMR), the latter of which primarily focused on science. The SOMER review made a number of recommendations for fundamental changes to management and operations, to improve flight-hour production and reduce costs. The FMR review suggested that transformative change was needed, with a strong need to alter planning and decision-making so that it is more science focused. The FMR gave a number of recommendations for completing high priority science programs and delivering high-quality data.

The survey committee found no evidence that SOFIA could, in fact, transition to a significantly more productive future. There have been only modest improvements in productivity over the past 2 years. These include a 50 percent increase in papers per year, and a higher completion percentage of high-priority programs. A new director was also recently appointed. It is noted, though, that the SOFIA team has responded to NASA that a number of major recommendations from the SOMER and FMR reviews are not feasible to implement, suggesting any future improvements would still be modest, and insufficient to bring about the flagship level science associated with its budget. Thus, the survey committee found no path by which SOFIA can significantly increase its scientific output or relevance to a degree that is commensurate with its cost.

Conclusion: The cost of SOFIA's yearly budget is comparable to NASA's Hubble and Chandra flagship missions, yet the scientific productivity is significantly lower. There is no evidence that SOFIA could transition to a significantly more productive scientific future.

Recommendation: NASA should end SOFIA operations by 2023, consistent with NASA's current plan.

6

Technology Foundations and Small and Medium Scale Sustaining Programs

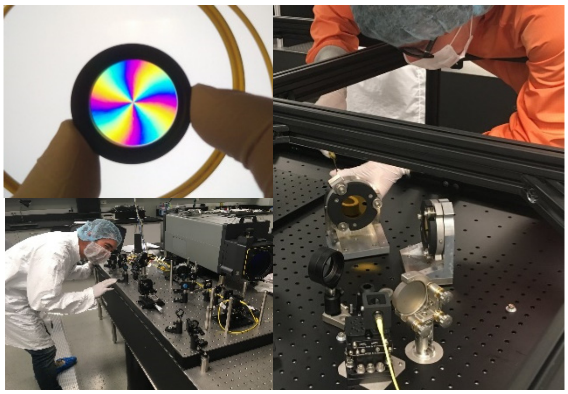
In this chapter the focus shifts from current facilities to the technology development that keeps them on the cutting edge, and the small and medium projects that complement them. These elements provide rapid response to new opportunities and discoveries, and offer platforms for building a strong and diverse community of innovative instrumentalists and technologists who will drive future progress. The agencies' historical willingness to support a significant range of program scales is a proven strength of the Nation's astrophysics portfolio, and is an even more pressing need today, as made clear by the large costs and long development timescales for the MREFC observatories and flagship missions submitted for consideration to Astro2020. This chapter draws from the Enabling Foundations panel report, as well as from the EOS-1, EOS-2, OIR, RMS and PAG studies, all of which emphasize the need for sustaining a broad range of activities for advancing Astro2020 science goals.

Small and mid-scale programs advance broad-reaching astrophysics scientific goals, and fuel new discovery. NASA's suborbital and Explorer missions, and NSF's Mid-Scale Innovations Program (MSIP) projects can be conceived, implemented, and deployed on a few-year timescale, in exchange for focusing on a narrower set of capabilities or science objectives. In addition to advancing a broad range of science simultaneously and synergistically, small and mid-scale projects are essential to the agility of the science program. Astrophysics is fundamentally a discovery-driven science, and examples of major advances enabled by the ability to respond quickly to new discoveries abound. The Swift Medium Class Explorer (MIDEX), with its agile pointing and broad wavelength coverage, was conceived, developed, and launched within 6 years of the discovery of the X-ray and optical afterglows of gamma-ray bursts. Another example is the Transiting Exoplanet Survey Satellite (TESS) Explorer mission, which was able to quickly capitalize and expand on the transit detection breakthroughs of Kepler to execute an all-sky census to identify potential James Webb Space Telescope (JWST) targets. On the ground, the DSA-110 MSIP radio array project was selected, developed, and is projected to be on-sky in the early 2020's rapidly responding to the progress in the field of Fast Radio Bursts (FRBs). None of these capabilities could have been met with a current or planned larger project, and astronomy's rapid response to these new scientific opportunities has been a proven success that we aim to replicate in the coming decade.

The range of institutions, both public and private, that engage in technology development, small missions and experiments, and mid-scale activities such as MSIP and Explorers is another major strength of the U.S. program. Collaborations involve public and private university-based efforts, government labs supported by DOE and NIST, and NASA centers. Industrial partnerships are also important at these scales, usually for development of components requiring specialized fabrication approaches or processes. Government laboratories and NASA centers house state-of-the art, sustained capabilities for, for example, metrology, lithography, and microfabrication, essential for many technical building blocks. Universities also have special expertise and dedicated laboratory facilities and testbeds that are often unique in the world. This combination has developed world-leading technology, and small and medium-scale observatories that have had high scientific impact for astrophysics (Box 6.1).

BOX 6.1 Development of Technology for Exoplanet Imaging and Spectroscopy: An NSF, NASA, and Private Partnership Working Toward a Grand Scientific Goal

Technology for imaging exoplanets has been developed in many labs, including Caltech's **Exoplanet Technology Laboratory** (ETL), UCSC's **Laboratory for Adaptive Optics**, the Princeton **High-Contrast Imaging Laboratory**, and others, combining NASA, NSF, state, and private funding.



Technology development at the ETL is undertaken by students and postdocs working side-by-side with experienced engineers.

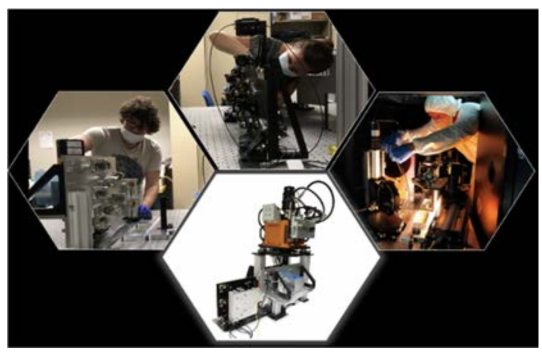
Connections from the Ground to Space

Technology from these programs has been incorporated into instruments such as the Gemini Planet Imager (GPI), Magellan MagAO-X, and the Keck Planet Imager and Characterizer (KPIC). These instruments use concepts developed in laboratories to study young giant planets (below left) while demonstrating techniques and technology that will someday measure the atmospheres of other Earths on a large UV-OIR space mission and the ground-based Extremely Large Telescopes.

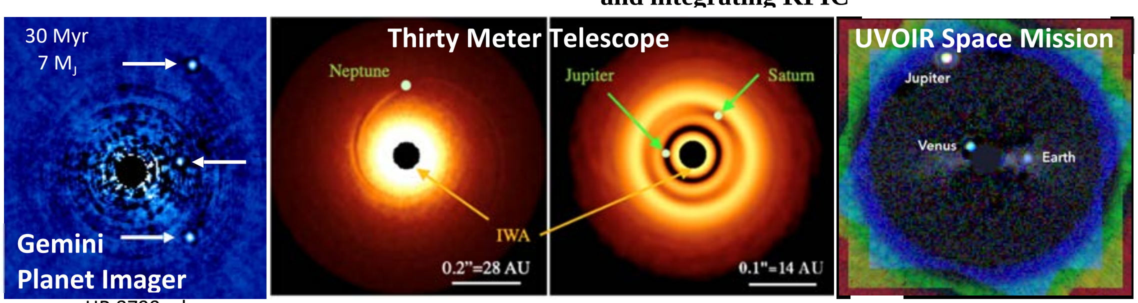
Starshade testbed at Princeton University

Technology, Testbeds, and People

Students, postdocs and test hardware are funded by **NASA's APRA** and **SAT**, **NSF's ATI** and **STC** programs, as well as **private foundations**. Example technologies tested in university laboratories include optical vortex coronagraphs, MEMS deformable mirrors, shaped-pupil coronagraphs, and new wavefront sensors. These laboratories have centrally involved graduate students and postdoctoral researchers who work with experienced engineers. Many of these young scientists are now in faculty or NASA staff positions establishing their own efforts.



Undergraduate and graduate students testing and integrating KPIC



HR 8799 cde

The Search for Habitable Exoplanets from Ground and Space

The ultimate challenge for high contrast imaging and spectroscopy is to search for atmospheric biomarkers indicative of life. This will be undertaken by the Extremely Large Telescope (above center) for small host stars, and a large UV-IR space mission (above right) for Sun-like stars. The technology developed in these laboratories will enable both efforts.

SOURCE: *Upper right:* Courtesy of Princeton University. *Middle left:* (from upper left, clockwise) Courtesy of Daniel Echeverri, Caltech Exoplanet Technology Laboratory; Courtesy of Jacques-Robert Delorme / Caltech Exoplanet Technology Laboratory; Courtesy of Jorge Llop-Sayson, Caltech Exoplanet Technology Laboratory. *Lower right:* People: Courtesy of N. Jovanovic et al., 2019, arXiv:1909.04541. Reproduced with permission. Instrument: Courtesy of Charlotte Z. Bond, et al., "Adaptive optics with an infrared pyramid wavefront sensor at Keck," Journal

of Astronomical Telescopes, Instruments, and Systems. 6(3) 039003 (24 September 2020) https://doi.org/10.1117/1.JATIS.6.3.039003. *Bottom row:* B1. Courtesy of the Gemini Planet Imager Exoplanet Survey Team. B2. Fitzgerald, M., et al. (2019). "The Planetary Systems Imager for TMT." Bulletin of the AAS, 51(7). Retrieved from https://baas.aas.org/pub/2020n7i251 Reproduced with permission. B3. R. Juanola Parramon/N. Zimmerman/A. Roberge (NASA GSFC).

Finally, these partnerships and the balance and range of project scales have been essential in developing the careers of the instrument builders, technologists, and PIs that are so important to the success of the astronomy and astrophysics enterprise. Being an effective PI of a large facility or flagship instrument, or a Small Explorer (SMEX) or MIDEX mission, requires a high degree of experience and training. These are often acquired through involvement with, and/or leadership of smaller payloads, or modest-sized ground-based instruments. The specialized training of technologists and instrument scientists is a progressive process, from the undergraduate level where students often first become engaged in the field, to graduate training in labs and on experiments, to early career stages where individuals develop their own initiatives and become established researchers. The endeavors undertaken during these career stages often progress with project size.

6.1 THE TECHNOLOGY FOUNDATIONS

New technologies for astronomical instrumentation are crucial building blocks without which observational capabilities would stagnate. It is hard to imagine modern astronomy without large-format CCDs, or without the bolometers and calorimeters that are at the heart of so many observatories and experiments, from time domain facilities to forefront Cosmic Microwave Background (CMB) polarization measurements. Early and significant investments in technology directed at flagship missions and large NSF facilities provide a refined understanding of costs and risks prior to construction, and for space missions it reduces the likelihood and magnitude of cost and schedule overruns during development.[1] NSF's Advanced Technology and Instrumentation (ATI) and NASA's Astrophysics Research and Analysis (APRA) and Strategic Astrophysics Technology (SAT) programs support several essential functions: supporting the modifications required to apply technologies to the exacting needs of astronomers; demonstrating that they function in the relevant environment and as part of a system; and inventing entirely new approaches for novel astrophysical measurements.

6.1.1 NASA's Competed Technology Development and Demonstration Programs

In this section we focus on early-stage technology development, investments in technology required to advance NASA's small and medium-scale missions, and maturation of component technologies to a level that they are ready to be incorporated into flight missions of all cost scales. We address the crucial issue of technology maturation for defined, strategic missions in Chapter 7. NASA supports two major technology programs for astrophysics—APRA and the more recently established SAT program—to support "blue-sky" and strategic, mission-oriented technology development, respectively.

6.1.1.1 APRA Technology Development

APRA's success is grounded in its open, competed calls for early-stage technology development as well as maturation and demonstration of component technologies.[2] The technologies developed through this program have advanced NASA's entire range of mission scales—from suborbital payloads,

[1] J. C. Mankins, "The critical role of advanced technology investments in preventing spaceflight program cost overruns," The Space Review, December 1, 2008, https://www.thespacereview.com/article/1262/1.

[2] APRA also supports the development of suborbital payloads and laboratory astrophysics. Here we are concerned with the subset of APRA supporting technology development.

to SmallSats, Explorers and Flagships over the full electromagnetic and multi-messenger spectrum, from submillimeters to gamma rays to cosmic rays. APRA is also the best opportunity in NASA Astrophysics for developing highly innovative but risky new technologies. APRA technology grants are also an important mechanism for early-career instrumentalists or technologists to establish themselves, and these grants fuel the university-based laboratory development efforts that train the next generation of innovators (Box 6.1).

The APRA technology funding is, however, significantly constrained. In addition to general technology development, APRA funds a wide range of activities, from suborbital payload development and science to laboratory astrophysics. It is therefore difficult to determine the exact amount supporting new technology; the Enabling Foundations Panel report estimates that 40 percent of APRA funding, or ~$8 million a year, goes into the Detector Development and Supporting Technologies components of the program (H.2.7.1). Of concern is the fact that a typical 3-year technology development grant in either of these two categories ranges from $200,000 to a maximum $400,000 a year, with only a few awards funded at the higher level. This funding is sufficient for student and postdoc support, but not for equipment purchases necessary to start a new lab or research effort, or, for involving commercial partners in fabrication of elements involving non-recurring engineering costs.

The limited APRA technology funding levels restrict its impact relative to the priorities of this survey in several important ways. First, levels are too small to address the need to advance broad technologies to acceptable levels (technical readiness level, or TRL, 5-6) for incorporation in future Explorers, suborbital and SmallSat missions. The APRA technology funding levels are also such that establishing a new laboratory effort is essentially impossible without significant supporting infrastructure provided by the host institution (i.e. leveraging an existing optics, electronics or detector lab, or using institutional start-up funds). This creates barriers to entry for young researchers or for researchers establishing new directions, and it limits the range of institutions that can effectively compete, reducing the overall diversity of participants. The Nancy Grace Roman Technology Fellowship program is intended to give early career researchers the opportunity to develop skills to lead flight instrumentation projects by providing funding to establish a laboratory and research group. The program is excellent, and has supported individuals who are now PIs on suborbital and satellite missions. However, the funding levels of $300,000 are small given that such efforts are typically multi-year, and require the purchase of significant equipment.

Recommendation: NASA should increase funding levels for the Detector Development and Supporting Technology components of the Astrophysics Research and Analysis Program. Priority should be placed on increasing grant sizes for larger efforts as well as increasing the overall funding in the technology elements of the program. The total increase needed to ensure a healthy selection rate and appropriate grant sizes is estimated to be about 50 percent above inflation.

6.1.1.2 The Strategic Astrophysics Technology Program

The SAT Program, initiated in response to a recommendation from *New Worlds, New Horizons* (NWNH), competes and selects projects aimed at maturing component technologies relevant to strategic flagship missions to the point they are demonstrated at a subsystem level and/or in a relevant environment (TRL 6). The first selections from the 2012 call responded to specific flagship technology development needs identified by NWNH. Examples of programs funded from the 2018 call include demonstration of wave front control for a future high-contrast exoplanet imaging mission such as the proposed HabEx or LUVOIR, high-resolution far-IR receivers for a mission such as the proposed Origins Space Telescope, and adjustable high-resolution X-ray optics, at the heart of the proposed Lynx flagship. While directed at flagships, some of these technologies have potential application on Explorer class missions. As a competed program open to the community, SAT draws from a large talent base at universities, NASA centers and government labs.

The SAT program is an important element in addressing the maturation of component technologies at the intermediate level (TRL 3-5), however it is insufficient to address the need to co-mature mission concepts and their associated technologies in a coherent way. Chapter 7 discusses this issue, and recommends establishing The Great Observatories Mission and Technology Maturation Program to address this gap. There will, however, still be the need to mature technologies for the Probe class missions, as well as for strategic missions prior to their funding through the Great Observatories Mission and Technology Maturation program

Recommendation: NASA should continue funding for the Strategic Astrophysics Technology Program, and should expand proposal calls to include intermediate level technology maturation targeted in strategic areas identified for the competed Probe class missions.

6.1.2 NSF's Advanced Technologies and Instrumentation Program

NSF's ATI program is a critical component of the AST portfolio that supports the development of innovative, potentially transformative technologies (even at high technical risk) within the overarching AST science objectives. Although ATI is within the AST division, there is a natural overlap with broader programs such as Major Research Instrumentation (MRI) and Faculty Early Career Development (CAREER) Program and some awards are co-funded. Technologies and instruments supported under ATI span the range from radio through optical and have included epoch of reionization receivers, very-long-baseline interferometry (VLBI), CMB experiments, Microwave Kinetic Induction Detectors (MKID), infrared (IR) detectors (Figure 6.1 shows an example), CCDs, adaptive optics, large mirrors, laser frequency combs, integral field units, specialized software, and more. In addition to technological advances, many awards lead to significant advances in observational capabilities. ATI has supported projects that are small enough to be managed by a single investigator, yet large enough to have a substantial impact.[3] Crucially, ATI funding is one of the few mechanisms through which an early-career instrumentalist can become established, and these projects provide essential training for students and postdocs.

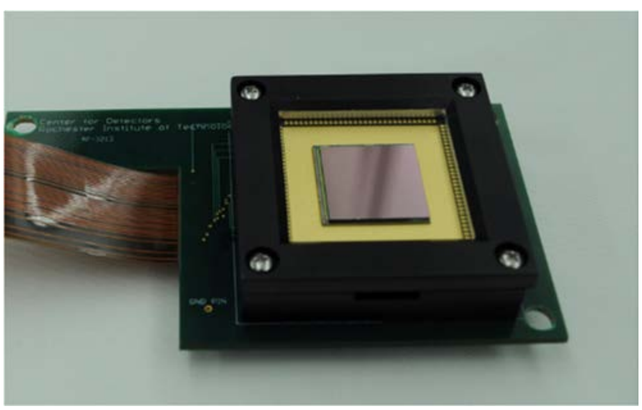


FIGURE 6.1 ATI-1. Low cost infrared detector arrays for space and ground. One of the goals of the research is to produce low-cost large-format devices (up to 8000 x 8000 pixels) for the next generation of ground and space-based telescopes. The research was supported by NASA's APRA program, NSF's ATI program, and made use of the NSF supported Materials Research Science and Engineering Center (MRSEC) facility at Cornell University.

[3] P. Kurczynski and S. Milojevic in "Enabling Discoveries: Thirty Years of Advanced Technologies and Instrumentation at the National Science Foundation," arXiv:2001.05899.

SOURCE: From Brandon J. Hanold et al., "Large format MBE HgCdTe on silicon detector development for astronomy", Proceedings of SPIE 9609, Infrared Sensors, Devices, and Applications V, 96090Y (28 August 2015); doi:10.1117/12.2195991

ATI addresses a crucial stage in instrument development. Putting major facilities aside, a modern millimeter-wave, IR, or optical instrument built for an existing ground-based telescope costs anywhere between a few and roughly $30 million, and then up to $5 million more to characterize, calibrate, operate, and deliver usable data. Bringing one of these to fruition requires careful attention to cost, schedule, and management. While there is room for some innovation, the technological foundation for these projects needs to be fairly solid for success. An underdeveloped technology can lead to delays and cost overruns in a large project. This risk can be mitigated by advancing technology through the ATI program.

Past decadal surveys have recommended increased investment in developing basic technology. Despite this advice, however, NSF AST has instead significantly cut the budget for ATI over the last 10 years. Astro2010 recommended an increase in ATI funding, from 10 million a year (FY2010) to 15 million a year to accommodate general technology development, including the pressing need to develop advanced adaptive optics systems in the optical, as well as new radio instrumentation. Instead, since 2012 the ATI budget has steadily decreased until today, where for the last 3 years it has been funded at the $6 million a year level (FY2020). This contraction has significantly restricted the size of awards, so that they are no longer sufficient to develop an advanced technology without requiring researchers to seek and juggle support from multiple sources, and to rely heavily on existing infrastructure. Such infrastructure may not be available to new researchers, especially those at institutions without established technology development efforts. Finally, ATI funding is insufficient for developing small scale instrumentation (less than a few million dollars), and so the NSF-wide MRI program is the only avenue available for this. However, MRI is highly over-subscribed, requires institutional matching, and has institutional limits on the number of proposals that can be submitted. These factors severely limits the opportunities available for astrophysics instrumentation.

Looking to the coming decade, the need to support advanced technologies is, if anything, greater than it was a decade ago. In Chapter 7 an expanded mid-scale program is recommended, the success of which will depend on novel technologies and approaches. The survey committee also recommends U.S. investment in very large telescopes. These will transform science, but not without state-of-the art instrumentation and adaptive optics (AO) systems that enable diffraction limited observations, which require significant technology investment. Other areas ripe for investment include, but by no means are limited to, correlators and elements of radio cameras, far-infrared detectors and spectrometers, predictive control for adaptive optics, ultraprecise radial velocity techniques, and advanced fiber positioning systems for massively multiplexed spectrographs. To ensure the future has a strong foundation in technology and instrumentation, ATI funding must be increased, a sentiment also supported by the report of the Panel on an Enabling Foundation for Research.

Recommendation: The National Science Foundation should restore the Advanced Technologies and Instrumentation Program to $14 million a year (fiscal year (FY) 2020)—the same level of support it had in 2010—and further increase it to a target level of $20 million a year (FY 2020) by 2028.

6.2 SMALL AND MEDIUM-SCALE PROGRAMS

As described above, small and mid-scale projects and missions are essential to sustaining scientific advances because of their speed and nimbleness in responding to new scientific opportunities, their ability to extend the wavelengths and techniques with which we observe the universe, their essential role in maturing new, transformative technologies, and their function as platforms for cultivating the next generation of instrumentalists and technologists who will build the facilities of the future. An analysis of

specific small and mid-scale instrumentation and mission programs at NASA and NSF is presented below.

6.2.1 NASA's Small and Medium-Sized Projects and Missions

Small- and mid-scale programs are absolutely essential for NASA. Not only are they key elements of a scientifically balanced portfolio, but they also address the fact that space missions are both more demanding and higher stakes than comparable ground-based projects. As such, NASA's smaller programs that demonstrate technology and develop skilled future PI's are core to the long-term success of its entire astrophysics portfolio. This section describes NASA's small and medium flight programs. These are all openly competed, with projects ranging in scale from a few million dollars to ~$300 million. All of these programs emphasize scientific return in the near- and long-term, provide opportunities for immediate science (on the timescale of a graduate student education), and build the foundation for future missions of all sizes.

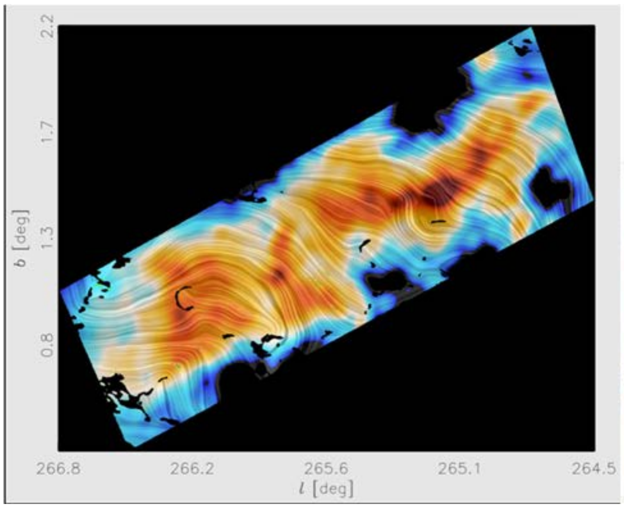
6.2.1.1 The Suborbital Programs

NASA's sub-orbital program addresses a wide variety of science, develops and tests essential technology for future missions, and trains the next generation of instrumentalists and project leaders. The sub-orbital portfolio comprises two components: high-altitude ballooning for reaching altitudes of up to 40 km for many days at a time (Figure 6.2) and sounding rockets to reach beyond the stratosphere for flight durations measured in minutes. Unlike orbital missions, sub-orbital programs allow rapid revision and reuse of payloads, speeding the technology development cycle. Many of NASA's largest visions have built upon the technology and expertise developed through these programs. Time and again, the program has demonstrated its efficacy in producing leaders for space missions. As touched on elsewhere in this report, cultivating new instrumentalists is essential for maintaining and diversifying both future leadership in space astrophysics overall, as well as ensuring a technically trained scientific workforce.

6.2.1.1.1 The Balloon Program

The balloon program offers access to a near-space environment with a wide variety of options for duration and sky coverage. Its wide array of capabilities include single day "conventional" flights and Long Duration Balloon (LDB) flights lasting up to 60 days in circumpolar flights around the Antarctic. After years of development, super-pressure balloons for Ultra-Long Duration Balloon (ULDB) flights carrying payloads up to 2000lbs with nearly constant float altitudes for up to 100 days, and are coming to the fore, opening up new possibilities. The balloon program's impact on innovation and science can be seen in its breadth of payload instrumentation: kilo-pixel IR and mm-wave cameras (Figure 6.1), CMB polarimeters, stabilized platforms with sub-arcsecond pointing accuracy for wide-field UVO imaging, gamma ray detectors, and sub-atomic particle detectors. Multiple proposed satellites with CMB, IR, X-ray, and time-domain capabilities submitted to Astro2020 have roots in the balloon program, just as their predecessors did for existing and completed explorers and flagships.

Pathways to improving the balloon program to take maximum advantage of these promising opportunities include: (1) increasing the number of flights; (2) continuing to strive for higher ULDB float altitudes; (3) increasing the accessibility of the program to more PIs by reducing barriers to entry; and (4) exploring structural adjustments that can support new PIs. Possibilities for accomplishing (4) include "piggy-backing" instruments on existing payloads, providing common hardware, providing access to funded engineering and mentoring support, and combinations of these.



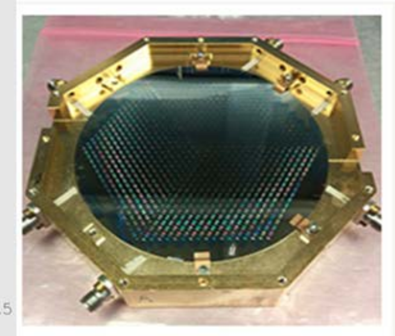


FIGURE 6.2 (*left*) Image of the Vela C molecular cloud taken by the balloon-borne BLASTPol instrument, showing the thermal emission at 500 microns with the direction of the magnetic field superimposed. The data provide new insights into the properties of dust and the role of magnetic fields in the interstellar medium through a wide range of densities. (*right*) Novel receivers based on Microwave Kinetic Inductance Detectors (MKIDs) are being developed and demonstrated as part of the BLAST program. SOURCE: *Left:* BLASTPol Collaboration/J.D. Soler. https://sites.northwestern.edu/blast/nearby-molecular-clouds. *Right:* B. Dober/NIST. See https://sites.northwestern.edu/blast/detectors.

Although it is beyond the scope of this survey to perform an in-depth analysis of the program, it is clear from the Enabling Foundations report (H.2.8.1) and the progress addressing NWNH recommendations that important challenges lie ahead for achieving the goals enumerated above, particularly for taking advantage of new ULDB opportunities. The first challenge relates to the available funding levels for balloon payloads, which have not kept up with the increased scope and complexity. Although NWNH recommended increasing the funding by $5 million a year to the R&A program (pg. 222, along with a $10 million a year increase for infrastructure), the budget has remained roughly constant over the decade at $25 million a year in awards typically supporting approximately 30 payloads in various stages of build, standby, deployment, and analysis. A second challenge relates to the ballooning infrastructure and management, which requires investment and possibly reorganization to find the right balance between increasing the launch rate and balloon technology development, while recognizing the inherent risks. Finally, broadening and diversifying participation will require changes to the way NASA supports teams, particularly those with young investigators at institutions that are still developing strong, independently funded technical and engineering infrastructure.

> **Recommendation: NASA should undertake an external review of the balloon program to establish a framework for accomplishing the competing needs of achieving flight capabilities and launch rates that meet demands, ensuring adequate investment in payloads, and lowering barriers to entry.**

6.2.1.1.2 The Sounding Rocket Program

Sounding rockets complement balloons by providing quick access to near-space conditions. This is a unique capability for some investigations, especially in wavebands where the residual atmosphere at balloon altitudes is limiting, such as soft X-ray, UV, and some infrared bands. Rockets are also crucial for maturing technologies and formally qualifying them for spaceflight. Because the pointing platforms are provided by NASA to the investigator teams, the barrier for entry is lower than for the balloon program, where groups typically must develop both the payload and pointing platform. This makes sounding rockets

attractive for developing new PIs, and diversifying instrument development teams. While there is limited proposal pressure from the community for increased flight rates and capabilities, the current sounding rocket program provides an important component of NASA's astrophysics program.

> **Conclusion:** The rocket program provides unique, irreplaceable opportunities for accessing space. It is important to maintain this capability.

6.2.1.1.3 NASA's Explorer Program

NASA's Explorer Program provides opportunities for competed, PI-led missions on a range of scales, from the SMEX and MIDEX missions with dedicated launches, to Missions of Opportunity (MOs), to the relatively new SmallSats. The stand-alone SMEX and MIDEX platforms, with PI-managed cost caps of $145 million (FY2020) and $290 million (FY2022), respectively, enable teams to propose highly capable but focused scientific missions. These are developed and launched on 5-year timescales, and respond to new discoveries while often providing multiwavelength capabilities distinct from those of NASA's flagships. MOs allow small payloads to be deployed on a variety of platforms, including ULDBs and instruments attached to the International Space Station (ISS). Recently, SmallSats—which include volume-limited CubeSats and other small orbital experiments launched as secondary payloads—have been added. The Enabling Foundation report (H.2.9) provides additional background, including recent selections.

From the small-scale MOs to MIDEXs, NASA's Explorer program has provided tremendous scientific return for the investment (Figure 6.3). Reflecting the program's value, NWNH recommended increasing NASA's investment in Explorers from $40 million to $100 million (FY2010) annually in order to increase the rate of proposal opportunities and launches. NASA has largely achieved the recommended target, and the resulting selections in the last decade, ranging from the TESS exoplanet mission to the Neutron Star Interior Composition Explorer (NICER) X-ray timing payload deployed on the ISS, have returned exceptional science during their prime mission phases, and have also served broad user communities in their extended missions.

> **Conclusion:** NASA's augmentation of the Explorer program in response to NWNH's recommendation has resulted in an increased rate and a tremendous science output.

> **Recommendation: NASA should maintain Explorer launch rates at the level specified in *New Worlds, New Horizons in Astronomy and Astrophysics*.**

The addition of SmallSats to the Explorer Program supports the development and launch of larger (12U) CubeSats and similar-scale satellites.[4] While not strictly sub-orbital, the SmallSat program has some common attributes to the balloon and sounding rocket programs. A SmallSat consists of both an instrument and a spacecraft bus that provides power, communications and pointing. The spacecraft can be commercially procured, meaning that, like the sounding rocket program and unlike the balloon program, teams can benefit from commercially provided infrastructure, and can focus on the instrument and science, potentially lowering the barrier to entry to new PIs and teams. Managing a SmallSat program can be challenging, and support provided by NASA could further increase the range of institutions participating in the program. It remains to be seen whether SmallSats will, in the long run, prove to be an effective platform for a range of astrophysics investigations. It will be important as it does with all elements of the program, it will be important for NASA to periodically review the proposal pressure and viability of SmallSat Explorer selections with the aim of achieving broad goals that include science

[4] CubeSats are spacecraft sized in units (U), each having a volume of about 10 cm x 10 cm x 10 cm. 1U, 3U, 6U, and 12U are common CubeSat sizes.

return, technology development and maturation, and broadening participation to advance diversity and inclusion.

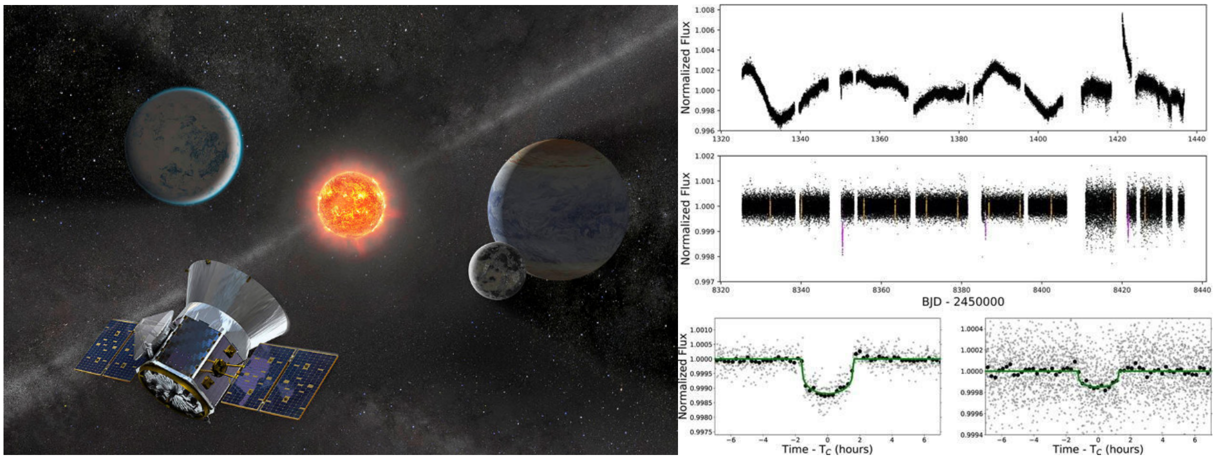


FIGURE 6.3 (*left*) NASA's Transiting Exoplanet Survey Satellite (TESS) Explorer mission which provides nearly continuous, high-cadence, ultra-precise optical photometry (light curves) has ushered in the era of exoplanet science and time-domain astrophysics on a large scale. Launched in 2018, during its prime mission TESS surveyed some 400,000 bright stars across the entire sky, with a cadence of 2 minutes and a typical duration of 1 month. TESS has already identified more than 4,000 planet candidates (more than 100 confirmed) and is ultimately expected to find 10,000 or more. The TESS GO program has also led to time-domain discoveries and followup ranging from near-Earth objects such as comets, to eruptions from active galactic nuclei (AGN) (in concert with NASA's Swift Explorer mission), to tidal disruption events caused by stars being disrupted by black holes. (*right*) TESS light curve of the K-dwarf star HD 21749, exhibiting transits by a sub-Neptune-size planet (2.6 R_earth) and an Earth-size planet (0.9 R_earth). SOURCE: *Left:* NASA TESS. *Right:* Adapted from D. Dragomir et al 2019, "TESS Delivers Its First Earth-sized Planet and a Warm Sub-Neptune," *The Astrophysical Journal Letters*, 875 L7. © AAS. Reproduced with permission. doi:10.3847/2041-8213/ab12ed

The highly scientifically successful Explorer program has challenges to overcome to address the lack of diversity in its scientific and technical teams. For the MIDEX and SMEX missions in particular, teams lack a healthy representation of career stage, gender, ethnicity, and institutional participation. Using the participation by women in mission leadership and science teams as one marker of diversity, one Astro2020 white paper finds that from 2008 – 2016 this participation was well below the representation of women in astronomy and astrophysics as a whole.[5] This means that the Explorer program is failing to benefit from the entire available talent base, and the broadest range of the community is not fully engaged in the unique opportunities presented by the program.

Effective leadership as PI for a SMEX or MIDEX scale mission requires significant experience and training. A first step to achieving a more diverse leadership pool is to broaden participation in technical, instrument, and leadership teams as a whole. However, especially for technical teams at small institutions, this is challenging due to structural barriers to entry. For instrumentalists and mission leaders (PIs, Project Scientists, and Instrument Leads), the complex, costly, and unique engineering and technical resources required to develop a mission proposal create significant barriers to entry. NASA, by design, does not compete the funding for mission concept or proposal development, leaving potential PIs to seek resources on their own. Teams with access to NASA or other specialized centers, and those with internal resources benefit overwhelmingly from this structure. While it is important not to limit potential proposers so as to have the largest range of concepts to choose from, provision of resources for mission

[5] J. Centrella, M. New and M. Thompson, 2019, Leadership and Participation in NASA's Explorer-Class Missions, white paper submitted to the Astro2020 decadal survey, https://arxiv.org/abs/1909.10314.

concept development by NASA through a simple proposal process would significantly lower the barrier to entry. It is also important that experienced PIs establish team roles that enable emerging leaders to gain experience. NASA's new Pioneers program (see below) is a potential stepping stone in this process. NASA's PI Launchpad Workshop held in 2019 is another welcome step in efforts to expand the range of future PIs. Additional suggestions are presented in Appendix H, the report of the Panel on an Enabling Foundation for Research, and NASA is also sponsoring a National Academies' study, "*Increasing Diversity and Inclusion in the Leadership of Competed Space Science Missions.*"[6]

> **Conclusion:** The NASA sponsored NAS study "Increasing Diversity and Inclusion in the Leadership of Competed Space Science Missions" will provide important advice towards broadening participation, and by implementing this advice NASA will strengthen the Explorer Program's overall success.

6.2.1.1.4 The Pioneers Program

NASA began the Pioneers Program in 2020 as a means of bridging the gap between stand-alone Explorer missions and suborbital platforms. This program has overlap with the balloon, rocket, and Explorer MO programs, however it is distinct in providing up to $20 million of funding, greater than that available for traditional suborbital and SmallSat platforms. It is specifically designed to provide opportunities for early-to-mid-career researchers to lead space or suborbital science investigations for the first time. Although still in its early stages, Pioneers are an important exploration of a path for broadening the pool for tomorrow's Explorer-class leaders while at the same time delivering important science. Three SmallSats and one balloon proposal, with goals ranging from measuring intergalactic UV emission to detecting ultra-high energy neutrinos, have been selected for further study from the inaugural announcement of opportunity. Encouraging the development of new leaders in space instrumentation and mission implementation is aligned with the Astro2020 objective to broaden participation in NASA's Explorer program.

6.3.2 NSF's Midscale Programs

NSF's competed midscale programs provide many of the essential elements advanced by NASA's suborbital and Explorer programs. NSF has two programs that support mid-scale projects, which for the purposes of this survey we define to be activities in a cost range from ~$4 million to ~$100 million, occupying the cost range between ATI grants and Major Facilities. One of these two programs is MSIP, funded and managed by AST since 2014. The other is the agency-wide Mid-Scale Research Infrastructure (MSRI) program founded in 2018 as part of NSF's 10 Big Ideas and managed by a cross-disciplinary team of NSF Program Directors. These multiple routes for funding midscale projects—MSIP and MSRI—have different funding streams: MSIP falls within the NSF AST budget while MSRI is NSF-wide. While this increases the diversity of funding opportunities, the total amount of funding available for astronomy and astrophysics projects faces uncertainties due to the added NSF-wide competition for the latter program.

The NWNH decadal survey recommended MSIP as its second highest priority for large programs on the ground. This competed program, based on NASA's highly successful Explorer model, was intended by NWNH to enable projects of size between the Midscale Research Implementation (MRI) program and less than typical for an MREFC project, or between ~$4 million and $135 million in the NWNH recommendation. NWNH also recommended calls for MSIP projects be open, peer reviewed, and

[6] https://www.nationalacademies.org/our-work/increasing-diversity-in-the-leadership-of-competed-space-missions

competed in two categories: conceptual design; and detailed design and construction. The total funding level for the program as envisioned by NWNH was ~$40 million a year (FY2010).

NSF implemented the recommended program leading to exciting, high-impact projects with broad science reach and relevance to Astro2020 science. In its first three cycles, MSIP has competitively awarded a total of $114 million to 18 distinct projects spanning a diverse range of science and wavelength (Figure 6.4). These awards have supported new projects, such as the Hydrogen Epoch of Reionization Array (HERA), designed to measure and characterize the universe from the cosmic dawn to the epoch of reionization, and the Deep Synoptic Array (DSA) that will pinpoint and study fast radio bursts.[7] MSIP has also funded upgrades and new instrumentation on existing telescopes, such as the Keck Planet Finder precision radial velocity instrument, as well as community access to existing facilities such as the Large Millimeter Telescope (LMT) and the Las Cumbres time domain optical follow-up network. The program has therefore provided broad access across public-private partnerships, has included international collaborations, and has advanced both individual-investigator initiated programs, large survey projects, and archival research.

However, MSIP has not approached the target total funding level set by NWNH, nor has it supported activities over the full range of cost scales, with most programs being at the lower end. To date, the selected awards have provided between $2 million and $12 million per project, significantly below the ~$100 million level envisioned by NWNH for at least some larger facilities. The last biannual solicitation provided a total of ~$21 million in funding, well below the $40 million a year target.

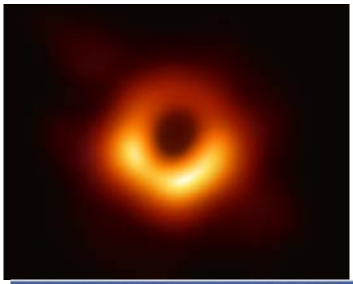
The Mid-Scale Research Infrastructure (MSRI) program funds a range of activities including facilities, equipment, instrumentation, or computational hardware or software. Divided into two tiers (MSRI-1 and -2), the most recent MSRI-1 call in late 2020 funded design and construction in the $6 million to $20 million range, and MSRI-2 funded infrastructure (construction) projects in the $20 million to $100 million range, excluding operations and science. In astronomy and astrophysics NSF has funded the design and development of CMB-S4, and design of the next-generation Event Horizon Telescope, both in the MSRI-1 category, and no MSRI-2 projects, with <~14 percent of the total agency-wide funding going to astrophysics projects. With the most recent solicitation, going forward this program will allow proposals across almost the entire mid-scale range envisioned by NWNH, a very welcome development. However, the oversubscription rate is extremely high, and an uncertain fraction will support AST and astronomy-related projects in the NSF Division of Physics (PHY). It is also not clear what the criteria are for preliminary selections, which are made by a panel of NSF program officers.

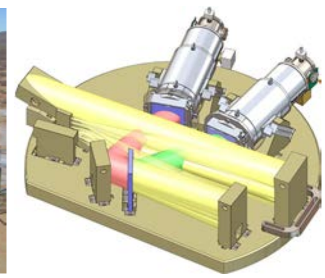
The survey received a large number of APC white papers for midscale projects that were evaluated by the OIR, PAG, and RMS program panels. Most of these were at the upper end ($50 million to $100 million) of the mid-scale range. While the survey did not request TRACE evaluations of any mid-scale concepts, the program panel studies were sufficient to determine that there is no shortage of compelling projects that could be accomplished with mid-scale funding. All three of the program panels that considered projects, as well as the Enabling Foundation panel that considered the program as a whole, strongly endorsed mid-scale projects, providing multiple superb examples of past accomplishments and compelling new mid-scale ideas. The panels all emphasize their high science value (H.2.10), cost effectiveness, and ability to enable agile approaches to addressing new science opportunities through the decade.

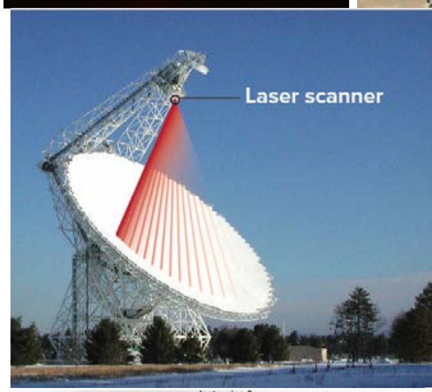
> **Conclusion:** Mid-scale programs across the entire range of scales (~$4 million to $100 million) are vital to the enabling foundation of astronomy research.

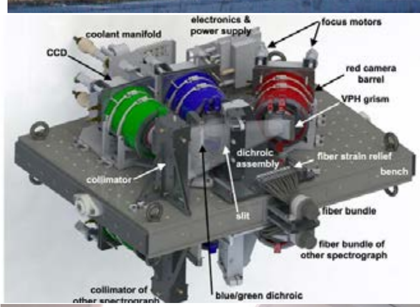
[7] NSF, "Award Search," accessed May 12, 2016, https://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=1257&BooleanElement=Any&BooleanRef=Any&ActiveAwards=true&#results.

As evidenced by the number of compelling community white papers, and given the assessments of the PAG, OIR, RMS and EF panels, the survey committee recommends in Chapter 7 expanding the mid-scale programs, including adding elements that ensure its responsiveness to decadal survey priorities.









Permission Pending

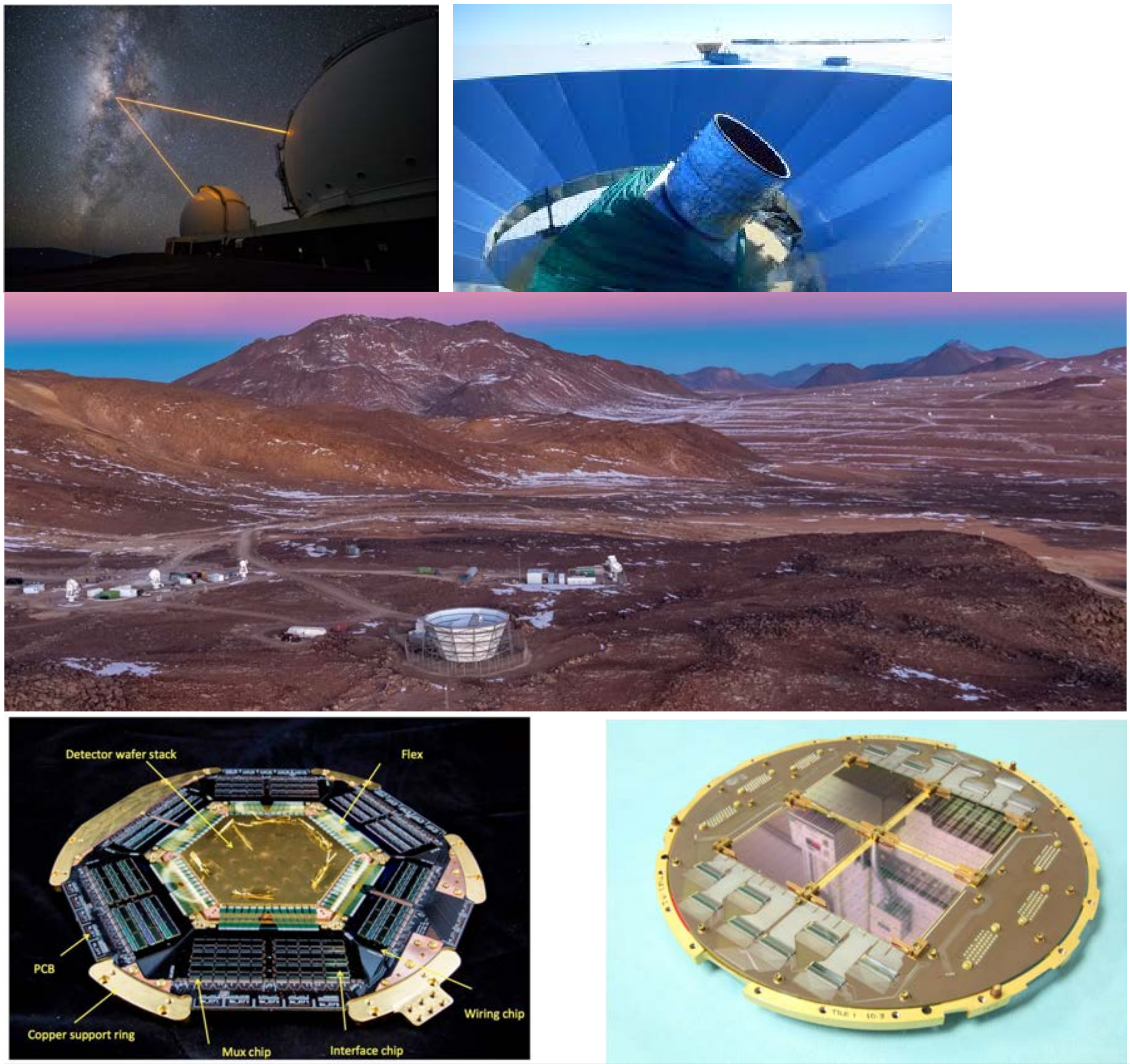


FIGURE 6.4 Hardware projects supported by the MSIP program between September 2014 and September 2021. *Left to right, from the top:* EHT, HERA, Keck Planet Finder, The Green Bank Telescope will use a laser scanning system to measure and adjust its surface precisely, CHARA Array's beam combining tables, LLAMAS Integral Field Unit, Evryscope and ARGUS array prototype, DSA-10 radio array prototype, MAPS: MMT AO exoPlanet Characterization System, LMT, and Keck adaptive optics (AO) systems. SOURCE: (1) The Event Horizon Telescope Collaboration. (2) HERA Partnership. (3) Keck Planet Finder, courtesy of California Institute of Technology. (4) Green Bank Observatory/Associated Universities, Inc. (5) Courtesy of Steve Golden/Center for High Angular Resolution Astronomy. (6) Adapted from Gabor Furesz, et al., "Status update of LLAMAS: a wide field-of-view visible passband IFU for the 6.5m Magellan telescopes," Proceedings of SPIE 11447, Ground-based and Airborne Instrumentation for Astronomy VIII, 114470A, 2020, doi:10.1117/12.2562803. (7) Courtesy of Nicholas Law and the Evryscope Collaboration. (8) DSA-10 radio array prototype. (9) Lori Harrison, Center for Astronomical Adaptive Optics, University of Arizona. (10) INAOE photo archive. (11) Courtesy of Sean Goebel Photography. (12) Adapted from L. Moncelsi et al., "Receiver development for BICEP Array, a next-generation CMB polarimeter at the South Pole," Proceedings of the SPIE 11453, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy X, 1145314, 2020, doi:10.1117/12.2561995. (13) Debra Kellner/Brian Bloss. (14) Courtesy of Yaqiong Li et al., "Assembly and integration process of the first high density detector array for the Atacama Cosmology Telescope," Proceedings of the SPIE 9914, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VIII, 991435, 2016, doi:10.1117/12.2233470. (15) The BICEP/Keck Collaboration, adapted from P.A.R. Ade et al., "BICEP2 II: Experiment and Three-Year Data Set," 2014, *Astrophysical Journal* 792: 62, doi:10.1088/0004-637X/792/1/62, © AAS, reproduced with permission.

7

Realizing the Opportunities: Medium- and Large-Scale Programs

The previous chapters laid out a roadmap for how multiple elements of the astrophysics enterprise need to work together to realize our scientific ambitions, and to do so in a manner that uses and supports the community's full potential. This chapter presents the decadal survey committee's recommendations for medium and large programs, program augmentations, and projects, as defined by their budgetary requirements.

The recommendations here flow from the science and program panel reports. The science panels identified a set of compelling and inspiring science questions that are organized into three broader themes in Chapter 2: *Worlds and Suns in Context*, *New Messengers and New Physics*, and *Cosmic Ecosystems*. These science themes reflect that we have entered a new phase of astronomical exploration in multiple dimensions, combining detailed characterization of known classes of objects with opening up the vast discovery space of the unknown. In addition to the broad science themes, the survey committee identified three priority areas that define the scientific frontiers, and motivate the recommended new investments in large projects: Unveiling the Drivers of Galaxy Growth; New Windows on the Dynamic Universe; and Pathways to Habitable Worlds. The steering committee aggregated and balanced the panel report and agency budget guidance inputs to arrive at the program described below.

We now see that the cosmos is dynamic, roiling and explosive with pulses of electromagnetic radiation, gravitational waves, and elementary particles streaming through space carrying messages of their exotic origins. Once separate lines of investigation, for example, black hole formation and large-scale structure, are now known to be inextricably intertwined. The ability to see biology's impact on the atmosphere—signatures of life—from distant exoplanets is also now within reach. With the same capabilities, we can also characterize the gaseous halos surrounding galaxies that fuel their growth and unravel how stars live, die, and seed the universe with the elements necessary for life. This vast array of phenomena is taking place in a universe filled with the cosmic microwave background whose study can tell us about transient sources, the contents of the universe, and the production of gravitational waves from the Big Bang.

In presenting the analysis of the highest-priority medium and large-scale projects to pursue, the survey emphasizes scientific breadth and balance of project scales. The interconnected science questions to be addressed require an interconnected program that recognizes and draws on program elements that vary in size, timescale, and wavelength/messenger, from radio waves to relativistic neutrinos. Astronomy is fortunate that most of its facilities are multipurpose and can simultaneously address multiple distinct science questions. Likewise, a major theme stressed by the program panels and the survey committee's own analysis is that because different wavelengths and messengers provide such essentially different and complementary views of the universe, a diversity of observational resources is needed to tackle the questions identified by the science panels.

7.1 CONTINUING PROGRAMS AND PROJECTS IN DEVELOPMENT

The new recommended medium- and large-scale activities build on missions and projects from prior surveys that have yet to begin scientific operation (Table 7.1). This survey assumes that these

compelling programs will be completed and sustained through their scientifically productive lifetimes. Ambitious and transformative large-scale efforts often take multiple decades to realize, and all of those scheduled for completion in the coming decade will provide important capabilities upon which the survey's scientific goals rely. Further, programs resulting from decadal recommendations, such as NASA's Explorers and NSF's Mid-Scale Innovations Program play an essential role in sustaining scientific breadth and ensuring timely response to new opportunities. These continued and future capabilities are essential underpinnings of the survey's new recommendations.

On the ground, the Vera Rubin Observatory, with science commencing in late 2023 or early 2024, will "conduct a deep survey over an enormous area of sky, and do it with a frequency that enables images of every part of the visible sky to be obtained every few nights."[1] Several of the survey's priority programs are designed to support follow-up of the Rubin Observatory's static and dynamic observations. The Daniel K. Inouye Solar Telescope (DKIST) will complete commissioning this year, and will begin to observe the Sun's fundamental magnetic and plasma processes to elucidate the role that magnetic fields and their interactions play in driving solar activity. The Mid-Scale Innovations Program, established as a result of Astro2010, is essential to the scientific balance of the Nation's ground-based investments and has proven to be extremely cost-effective. While it has yet to ramp up to its envisioned level, it is already providing diverse scientific capabilities and community access to private facilities strongly emphasized by this survey. All three programs are essential to the scientific future, and further augmentations to the NSF mid-scale programs is one of the Sustaining Program recommendations (Section 7.6.2).

In space, the James Webb Space Telescope (JWST) is a powerful strategic mission expected to launch by the end of this year, that, among many other things, will reach back in time to observe the first stages in galaxy formation, complementing the survey's focus on unveiling the drivers of galaxy growth more locally. JWST will also characterize the inner parts of other solar systems and the potentially habitable worlds orbiting small M stars, laying the foundation for the Astro2020 program that will extend this to further distances and Sun-like stars. In the middle part of the decade the Nancy Grace Roman Space Telescope will begin its cosmology and exoplanet microlensing surveys, and with a field of view more than one hundred times greater than Hubble, will provide powerful new capabilities for General Observers (GO). NASA is also a partner in The European Space Agency (ESA)'s M-class Euclid mission, as well as the L-class Athena and Laser Interferometer Space Antenna (LISA) missions. Euclid will complement the cosmological surveys from Roman and Rubin. Athena realizes some of the capabilities of the International X-ray Observatory (IXO) recommended by Astro2010, and will probe the hot, energetic universe, and will make important and unique contributions to the Cosmic Ecosystems theme. LISA will expand gravitational wave sensitivity to low frequencies, and will be an important foundational component of the *New Windows on the Dynamic Universe* priority science area. These missions will provide unique and powerful observational capabilities and science reach not duplicated by any other elements of the program advanced by this survey. Finally, the Explorer program has recently reached the enhanced selection rates envisioned by Astro2010, and is providing high value scientific returns responsive to the emphasis on scientific breadth and balance of mission scales.

> **Conclusion:** The decadal survey committee's recommendations for advancing the new programs or augmentations are predicated on the assumption that the major astrophysics facilities and missions in NASA, NSF, and DOE's current plans are completed and fully supported for baseline operations and science. New recommendations for space are additionally predicated on the assumption that NASA's Explorer program maintains the current, healthy selection rate.

Section 7.7 provides some advice and recommendations for NASA regarding Roman, Athena and LISA intended to ensure optimal scientific return to the U.S. community from these important missions.

[1] See https://www.lsst.org/about

TABLE 7.1 Medium and Large Programs and Projects in Development

GROUND		
Mid-scale Innovations Program	Competed projects in the range $4 million– $100 million	Scientifically broad, includes new observing capabilities, instrumentation, archiving, and data accessibility.
Daniel K. Inouye Solar Telescope	4 m optical telescope	Stellar magnetic fields, Solar activity
Vera Rubin Observatory	8.4 m aperture optical wide field telescope	Nature of dark matter, dark energy, cataloging the Solar system, exploring the dynamic sky
SPACE		
Explorer Program	Includes completing current selections and maintaining a cadence of two MIDEX, two SMEX and four MoOs/decade.	Scientifically broad, including recent missions and MoOs emphasizing X-ray spectroscopy, polarimetry, all-sky infrared spectroscopy, and time domain astrophysics.
James Webb Space Telescope	6.5 m IR telescope	First galaxies, star and planet formation, cosmic feedback
Roman Space Telescope	2.4 m wide field of view O/IR telescope	Cosmology, exoplanet microlensing, GO program
U.S. contribution to Euclid (ESA led M-class mission)	1.2 m aperture telescope, optical imager/near-IR spectrometer/photometer	Dark matter, dark energy, expansion history of the universe.
U.S. contribution to Athena (ESA led L-class mission)	Large area X-ray spectroscopy and imaging	The hot and energetic universe: black holes, galaxy halos, neutron stars
U.S. contribution to LISA (ESA led L-class mission)	Low-frequency gravitational wave interferometer	massive black hole mergers, white dwarf binaries, stochastic background

7.2 BUDGETARY GUIDANCE FOR NEW PROGRAMS AND PROJECTS

A primary objective of the survey is to develop an integrated roadmap for astrophysics that is both aspirational and foundational, while also conforming to reasonable expectations regarding budgets available from the federal government. NASA and NSF presented budget guidance, which provided the framework within which the survey committee established its recommended program.[2]

NASA provided information on its Astrophysics budget in July 2019, which was then updated in August 2020. The budget was broken into two parts, NASA budgetary requests for FY2021 through FY2025, and a budget extrapolation from FY2026 through FY2040. This budget extrapolation started at $1.9 billion in FY2025. The most optimistic scenario provided by NASA, adopted as the guidance for this survey, has the budget growing after FY2025 at approximately 2 percent per year to $2.5 billion in FY2040 and beyond. For inflating project budget profiles, the analysis uses an inflation rate of 2.7% per year, as derived from NASA's new start inflation index for FY2020.[3] NASA's guidance for the most optimistic budget growth, shown by the solid blue line in Figure 7.10, is less than projected for inflation. Approximately $1 billion is held aside to fund existing NASA Astrophysics Programs; the remainder is available for strategic initiatives associated with the survey recommendations.

[2] See Section 7.8 for additional details of the NASA and NSF budgetary analysis.

[3]https://www.nasa.gov/sites/default/files/atoms/files/2019_nasa_new_start_inflation_index_for_fy20_final2.xlsx).

The NSF budget projections were presented by the agency in July 2019 and were updated in August 2020. The budget guidance was divided into two lines: the Astronomy Division (AST) budget, which covers ongoing research grants, education, facility operations, and administration; and the agency-wide Major Research Equipment and Facilities Construction (MREFC) line, which funds construction of medium to large facilities with costs exceeding approximately $135 million. The MREFC guidance ended in 2030, and it is extrapolated forward from this assuming 2.7 percent inflation, the same as that used for inflating project budget requirements.[4] Figure 7.8 (solid blue line) shows the MREFC budget profile consistent with the agency guidance. The NSF AST budget was given to the survey for 2019 only. Figure 7.9 shows the starting point, at approximately $290 million, consistent with the current AST budget and programs. The future growth required to accommodate inflation, as well as the initiatives recommended by the survey is shown in Figure 7.9, broken down by major constituent components.

The survey did not receive specific budgetary guidance from the DOE, however DOE communicated programmatic support for initiatives aligned with its scientific objectives, including investigation and refinement of fundamental cosmological parameters and the nature of dark matter, as aligned with the objectives of its High Energy Physics (HEP) program.

7.3 PRIORITIZATION PRINCIPLES

Multifaceted considerations are involved in identifying the highest priority programs and projects for investment of federal funding. Major investments must advance a bold and broad scientific vision, while at the same time ensuring a balanced program that responds to scientific opportunity. Astronomy and astrophysics advances in a global context, and the survey committee recognized and responded to the need for synergy with, and complementarity to, activities world-wide. Especially for ground-based observatories, private institutions and philanthropic entities have been, and continue to be central to some of the most ambitious endeavors. The survey committee carefully considered how to best leverage these private-public partnerships in a way that achieves ambitious science and also advances the aspirations of the entire community. There is also the challenging issue of balancing scientific ambition with feasibility and timeliness. All of these factors shaped the recommended programs, and their phasing.

Maintaining a program that is balanced along many axes was a driving objective in the design of the portfolio of projects and programs advanced here. Of primary importance is that this portfolio must be scientifically balanced and broad. Supporting a range of project scales is not only essential for achieving this, but also for balancing science return in the near term with longer-term, more ambitious goals. It also maintains the crucial flexibility to respond rapidly to new discoveries. The survey committee also strived to prioritize both efforts that would maximize return on new facilities that will come online in this decade, as well as projects that will bring about wholly new capabilities in the future. A final consideration was the need to build foundations for future large strategic missions so that their scientific capabilities and scope can be better understood in the next decade, while at the same time advancing the highest priority project now.

In prioritizing the new, large projects, the survey committee adopted a set of guiding principles. Primary among these is that large strategic missions and MREFC-scale observatories must each advance a broad set of Astro2020's priority science questions. The survey committee was also guided by the judgement that the estimated time from inception to science for any recommended project is an important consideration, and it must be based on a schedule analysis and assume optimal, but realistically achievable budget profiles. The rationale for this is to balance timeliness and scope in a rapidly evolving scientific landscape. These principles shaped not only the elements in the program advanced here, but also their scale. These guidelines also led to the identification of key decision points, designed to enable rescoping to adapt to project uncertainties and risks, as well as to uncertainties in the future budgetary landscape. These same guidelines can also be used to take advantage of opportunities if the budgetary

[4] Ibid.

climate improves due to changes in federal funding, private philanthropy, or increased international participation.

7.4 SUMMARY OF RECOMMENDED NEW MEDIUM AND LARGE PROJECTS AND ACTIVITIES (2023 – 2033)

The survey committee has developed a vision that capitalizes on the tremendous momentum and scientific opportunities before us this decade. Laid out below is an ambitious roadmap for high-priority space- and ground-based large and medium scale initiatives that are both compelling and ready to begin implementation in the decade 2023–2033 (Table 7.2). The program includes new, large-scale projects, as well as initiatives that sustain and build on past investments, that harness and advance the creativity of the entire talent base, and that prepare groundwork for future decades. The initiatives in this roadmap are given priority, based on criteria described above, over the many exciting projects that were presented by the community.

This program necessarily extends beyond the decade. At the large scale, the ambitious, transformative projects that the survey advances will take more than a decade to bring to fruition. In developing the program, key milestones are established where evaluation of these projects and guidance on their scope and direction will be required. These future assessments are needed to assure that the appropriate balance between ambition and timeliness is maintained considering evolving budgetary and technical realities. Cognizant that new opportunities will certainly arise in the coming decade in response to scientific discovery and technical advances, the survey balances present ambition with leaving room for future initiatives to start in the 2033–2043 timeframe.

The decadal survey roadmap advances the Astro2020 scientific agenda through a balance of major programs, projects and medium-scale observatories and missions. The priority activities are organized separately for ground and space, and for each they are binned into three categories: 1) sustaining programs; projects and programs that optimize science return from facilities coming on line in the decade, and that maintain scientific and program balance, 2) enabling programs that prepare for future large observatories, and 3) frontier projects; large strategic missions in space and MREFC-scale observatories on the ground. Specifically, the third Frontier Projects category includes space missions in excess of $1.5 billion, and those ground-based projects individually exceeding $135 million. Previous surveys have chosen to organize activities based on small, medium and large cost bins. In FY2020 dollars, medium scale projects for NSF would range from $30 million to $135 million, and for NASA, from $300 million to $1.5 billion. For this survey, the organization of the medium and large scale projects emphasizes the role the activity plays. The categories enumerated above are equally important to advancing the survey’s scientific goals. While they roughly correspond to cost scale, this mapping is not exact. Previous surveys have not prioritized projects in one category against projects in a different category, but rather emphasize the need for a balance programs at all cost scales. Similarly, this survey emphasizes scientific and program balance, and does not prioritize one category over the others. Within each category, there are instances where one activity has priority over others in the same category, due either to its scientific importance, or due to technical readiness and/or programmatic urgency.

TABLE 7.2 New Medium and Large Projects, Activities and Augmentations (2023–2033)

The Frontiers: Major New Projects (Space)

IR/O/UV Large Strategic Mission

- IR/O/UV telescope for exoplanet characterization and general astronomy. Mission-specific funding to begin mid-late decade after mission and technology maturation program
- Total implementation and operations cost (5 years) estimated at $11 billion[a]

Enabling Programs (Space)

Great Observatories Mission and Technology Maturation Program

Program to co-mature large strategic missions and technologies. First entrant: IR/O/UV observatory, Far-IR and high resolution X-ray observatories recommended to enter in second half of the decade

Sustaining Programs (Space)

Time-Domain Program (highest priority)

- A program of competed missions and missions of opportunity to realize and sustain the suite of capabilities required to study transient phenomena and follow-up multi-messenger events.
- Notional cost: $500 million–$800 million over the decade

Probe Line

- Competed line of cost-capped probe missions to bridge the gap between Explorers and strategic missions; focused on gaps in science and wavelength capabilities– this decade Far-IR and an X-ray complement to Athena
- $1.5 billion/mission, cadence of approx. one/decade

The Frontiers: Major New Projects (Ground)

Extremely Large Telescope (ELT) Program (highest priority)

- Federal investment in the U.S.-ELT program for the U.S. community
- $1.7 billion NSF share of $5.1 billion project

CMB-S4

- Stage 4 Cosmic Microwave Background Observatory
- NSF share $273 million, DOE share $387 million

The ngVLA

- Design, cost trade studies and prototyping to prepare for construction, which could begin by the end of the decade
- $2.5 billion NSF share of $3.2 billion project

Sustaining Projects (Ground)

Mid-scale Augmentation: Open, Strategic and Sustaining Instrumentation Calls

Augmentation to mid-scale programs; inclusion of *open calls* that tap into the creativity of the entire community, strategic calls that maintain scientific breadth and foundational instrumentation capacity, and sustaining instrumentation calls to optimize the return on the investment in existing facilities. Strategic calls for the coming decade are in time domain astrophysics (highest priority), radio instrumentation, and highly-multiplexed optical spectroscopy.

Programs and Facilities Funded through NSF Physics

Technology Development for Future Gravitational Wave Detectors

Technology development for LIGO upgrades, such as Voyager, and for next generation observatories such as Cosmic Explorer.

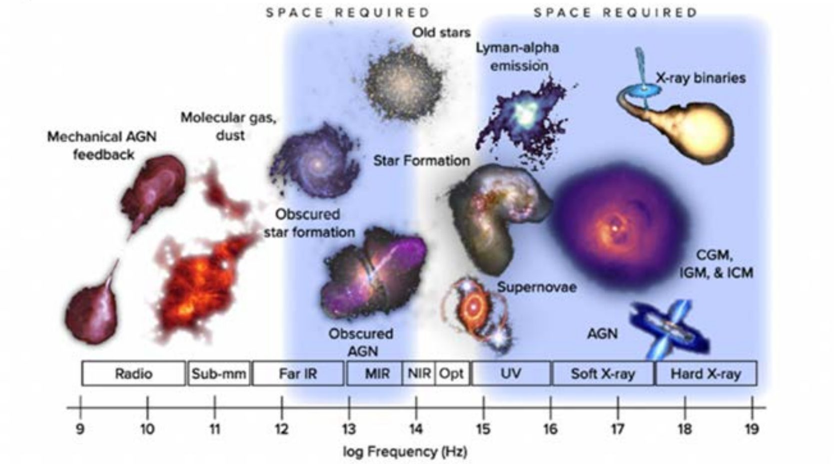
IceCube-Gen2 Neutrino Detector

Upgrades to the IceCube high-energy neutrino detector

[a] Project costs are in FY20$.

7.5 RECOMMENDED NEW ACTIVITIES FOR SPACE

The sections below describe the new space-based activities that the survey recommends NASA undertake in the next decade. In considering large strategic missions required to address the *Pathways to Habitable Worlds, New Windows on the Dynamic Universe*, and *Unveiling the Drivers of Galaxy Growth* priority areas, this survey takes a new approach. Rather than recommending missions in a rank-ordered list as they were presented in concept studies and white papers, the survey recommends a new strategy for rephasing mission and technology maturation and decadal survey recommendations (Section 7.5.1). This strategy is aimed at balancing the scope of large strategic missions to better address priority science in a timely way. Combined with this, a cost capped ($1.5 billion per mission) Probe line (Section 7.5.3.2), competed in strategic areas, will broaden the range of observational capabilities targeted at decadal science questions in the Ecosystems and Suns and Worlds in Context themes. Finally, a time domain program of competed small missions responds to the diverse observational needs required to address questions in the broad *New Messenger, New Physics* theme.



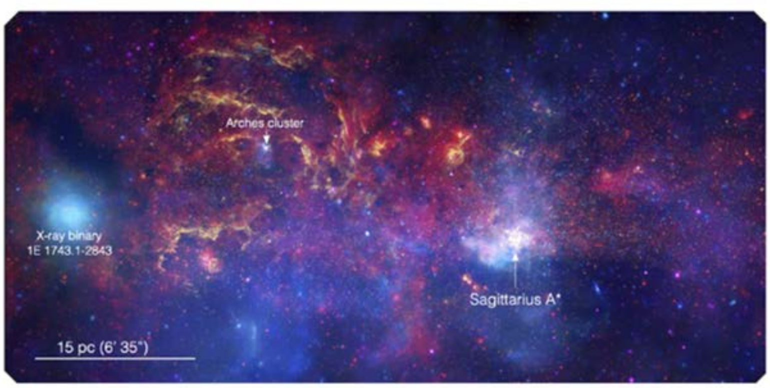


FIGURE 7.1 (*top panel*) Understanding the nature of the complex ecosystems operating inside galaxies requires observations across the electromagnetic spectrum. Many phenomena are only observable from space platforms (indicated by the blue shaded regions), and require large aperture, sensitive telescopes only achievable on a probe or large strategic mission scale. (*bottom panel*) The need for a nearly panchromatic suite of facilities is illustrated by the composite image of our own Milky Way galaxy made by NASA's Great Observatories: Chandra (X-ray), Spitzer (IR) and Hubble (UV/O). SOURCES: NASA, https://arxiv.org/pdf/2104.00023.pdf. NASA/Great Observatories Science Analysis Group.

7.5.1 Advancing NASA's Large Strategic Missions—the Great Observatories Mission and Technology Maturation Program

7.5.1.1 Background

The richness of the Astro2020 science calls for a broad range of observational capabilities spanning the electromagnetic spectrum. The power of broad wavelength coverage was demonstrated by NASA's Great Observatories, a panchromatic suite of four missions, launched over the course of three decades, that operated with contemporaneous overlap (Figure 7.1, Table 7.3).[5] These missions spanned a range of cost scales, from Hubble at approximately $10 billion in today's dollars, to Spitzer and the Compton Gamma-ray Observatory that today would be considered Probe scale. This was an extremely successful model that propelled major scientific advances on broad fronts over the course of two decades or more. The survey committee believes that it is scientifically compelling to replicate this approach today.

TABLE 7.3 NASA Flagship Mission Cost at Launch and Timescales

Observatory	Waveband	Launch date	Development (year)	2020 Cost (B$)
Hubble	UV/O	1990	18	9.4[a]
Compton GRO	Gamma-ray	1991	14	1.2
Chandra	X-ray	1999	17	3.1
Spitzer	IR	2003	11	1.0
JWST	IR/O	2021 (expected)	21	11
Roman	IR/O	2026 (expected)	14	3.5

[a] Cost at launch. Does not include servicing missions
NOTE: Shaded boxes indicate NASA's Great Observatory suite. Development timescale indicates time from survey recommendation to launch.
SOURCE: Data from https://arxiv.org/pdf/2104.00023.pdf.

While NASA's strategic missions must be driven by transformative scientific visions, they must at the same time advance a broad range of scientific objectives. If the scientific balance and wavelength breadth of the Great Observatories is to be replicated, these requirements must be balanced by the need for missions to be achievable on acceptable timescales. Balance and scientific breadth cannot be maintained if implementation costs are so large, and technology development so challenging, that projects can only be developed serially, with multiple decades required between inception of a mission concept by the community, adoption by a decadal survey, and the ultimate launch. These considerations are important in assessing how to best advance future large strategic astrophysics missions.

In preparation for Astro2020, NASA sponsored four Large Mission Concept Studies aimed at developing reference missions for consideration by the survey: Habitable Exoplanet Observatory (HabEx), Large UV/Optical/IR Surveyor (LUVOIR), Origins Space Telescope (Origins), and Lynx. These mission concepts were chosen by NASA as a result of broad community engagement. Over the course of several years, four Science and Technology Definition Teams (STDTs), working with a designated NASA center, developed the scientific case and possible mission implementation architectures, including required instrument capabilities. NASA then assembled a Large Mission Concept Independent Assessment Team (LCIT) to conduct technical, risk, and cost assessments that were

[5] Great Observatories the Past and Future of Panchromatic Astrophysics, 2020, https://arxiv.org/pdf/2104.00023.pdf.

independent of the STDTs.[6] This preparation and the mission definition studies were significantly more coordinated and uniform than has been done for previous decadal surveys. This advanced preparation provided a basis for the survey to evaluate the designs, performance, and likely budget scenarios.

TABLE 7.4 Large Mission Cost Estimates and Development Timescales

Large Strategic Mission	Waveband	LCIT Cost Bin (FY2020, B$)	TRACE Cost Estimate (FY2020, B$)	Development Time[a] (years, TRACE est.)
LUVOIR-B	UV/O/IR	8-10	17	20
HabEx 4-H	UV/O/IR	8-10	10.5	18
HabEx 3.2S	UV/O/IR	--	7.8	17
Lynx	X-ray	6 – 8	9	18.5
Origins	Far-IR	6 - 8	10.6	15.5

[a] Minimum, assuming immediate start and optimum budget profile.

The large mission concepts were studied in detail by the Panels on Electromagnetic Observations from Space (EOS-1 and EOS-2), which considered the mission science and evaluated its relevance to the key science questions and discovery areas identified by the Science Panels. In addition, the EOS Program Panels performed an assessment of the mission technology readiness, risk, and costs weighed against the scientific opportunities. Independent Technical Risk and Cost Evaluation (TRACE) studies performed by the Aerospace Corporation for five of the implementations (HabEx and LUVOIR both presented more than one architecture for consideration) were additional important inputs to the panels and steering committee. Table 7.4 summarizes the cost bins as determined by NASA's LCIT study, along with the TRACE estimates.[7] The LCIT binned missions into cost categories, and although lacking the extensive model-based analysis of TRACE, serves to bound the expected costs at the low end.

As noted above, establishing missions with complementary, panchromatic coverage operating near-contemporaneously, as done with the Great Observatories, can only be achieved if the missions in the suite can be launched in rapid enough succession, and have sufficiently long operating lifetimes to have overlap. While the mission concepts in Table 7.4 are all long-lived, their 15 to 20-year development times and large costs preclude more than one mission operating simultaneously, if they are built in a sequential development-then-launch cycle. The likely development times are in fact even longer than in Table 7.4, given that these assume the most optimistic funding profiles, which are rarely achieved in practice. The survey committee and EOS panels agree on this challenge, and take a two-pronged approach to addressing it by considering descopes to speed the development time, combined with a change in the development paradigm to allow some degree of parallel development to shrink the interval between launches.

> **Conclusion:** Establishing a panchromatic suite of observatories over the next 30 years is essential to address key questions in all three of the survey's priority science themes. The large strategic mission implementations presented to the survey cannot all be built and launched in an optimal timeframe given the current designs, available budgets, and approaches to mission development.

The universally long development times for the missions in Table 7.4 indicate that general purpose observatories with the full capabilities envisioned by the community will each take 15-20 years

[6] Large Mission Concept Independent Assessment Team (LCIT), 2019, https://science.nasa.gov/science-pink/s3fs-public/atoms/files/LCIT_FinalRpt-2019-Nov8.pdf

[7] See Appendix O for a discussion of the TRACE process.

to be developed and launched. The survey's recommendation to implement a probe mission line (Section 7.5.3.2) can partially address the need for scientific balance, however, to be consistent with the cost cap, probe missions must be significantly more focused than strategic missions. Accelerating the cadence of missions will require a combination of limiting their scope through careful selection of capabilities, and developing a new approach to the phasing of mission and technology maturation. This was done to some extent with the Great Observatories, in that both Spitzer and Chandra were significantly rescoped relative to the original concepts presented to the decadal surveys, yet both provided transformative observational advances.[8]

A rephasing of the mission and technology maturation process, with more significant and coordinated investment prior to a decadal survey recommendation to proceed with mission development, would provide multiple important benefits. This rephasing would recognize the multi-decadal timescales associated with large strategic mission and their associated technology maturation, and would better avoid the negative consequences associated with commencing missions prior to this maturation.[9] By investing more in the maturation process, NASA could develop missions to a level where there is significantly more confidence in the costs and requisite cost profiles before seeking Congressional approval for the final implementation. This would build a higher degree of confidence with the stakeholders. It would also make it more likely that the optimal cost profile could be obtained, speeding up development times and reducing the eventual total mission cost.

> **Finding:** For a decadal survey to confidently recommend implementation of a strategic mission as its highest priority, the mission's technology and architecture needs to be developed to a level of maturity that allows a reasonable assessment of budget profile, scientific performance, and technology risk. The mission's cost range and development timescale must be deemed appropriate for the scientific scope.

This survey suggests a restructured process for strategic mission maturation and decadal approval, shown schematically in Figure 7.2. The first stage is an initial mission **concept study**, similar to those done in preparation for this survey. However, rather than develop a small number of design reference missions, these studies would emphasize identifying key science break points, architecture options and trades, and would provide a description of key technologies and their maturation requirements and risks.[10] The **decadal survey** would then evaluate the missions' science promise and importance, and also, where possible, define the cost box and associated timescale deemed appropriate for the given mission. Based on the relative priorities, the survey would then recommend those candidate missions it finds to be compelling for significant investment in co-maturation of the mission design, science definition, and technologies through the **Great Observatories Mission and Technology Maturation Program.** With direction from an associated Program Office, these investments would take place in a phased way that represents the survey's priorities, and this phasing could be reevaluated by the mid-decadal to reflect circumstances (international landscape, technology progress, etc.). An **External Review**, either by a mid-decadal or decadal survey, or some other process external to NASA's usual program reviews, would decide whether the mission science capabilities and programmatic implementation is consistent with the decadal evaluation.

A process such as the one described above would address a number of important issues. If the initial concept studies clearly identify scientific, implementation and cost breakpoints, as well as technology development needs, the survey can better evaluate and recommend the appropriate scope for

[8] https://arxiv.org/pdf/2104.00023.pdf

[9] R. E. Bitten, S. A. Shinn, and D. L. Emmons, 2019, Challenges and Potential Solutions to Develop and Fund NASA Flagship Missions, NASA Technical Reports, https://ntrs.nasa.gov/citations/20190001455.

[10] Note: this was in essence what was done for HabEx and LUVOIR, which combined presented a number of point designs of differing scale all aimed at the goal of high contrast imaging and spectroscopy of extrasolar planets, with varying capabilities for additional astrophysics.

the mission in light of the survey's priority science. By recommending one or more missions enter into the maturation program in the coming decade, significant investments can immediately begin to refine them, develop their technologies and define the best mission possible for the recommended level of investment. Subsequent decisions about whether to start implementation of a given project can be based on the success and speed at which each project develops, and on the available budget. The survey committee expects that this process will result in decreased cost and risk and enable more frequent launches of flagship missions, even if it does require significantly more upfront investment prior to a decadal recommendation regarding implementation. This investment will not be wasted, even if a decision is made not to advance a particular mission in the subsequent decade. Many aspects of the technology investments, particularly in the areas of optics and sensors, will also be relevant for Probe and Explorer scale missions, and will benefit these programs.

> **Conclusion:** Enabling subsequent decadal surveys to recommend mission implementations with sufficient knowledge of the feasibility, overall budgetary needs, and timescale requires significant investment towards maturing large strategic mission science, technologies, and architecture in an integrated way.

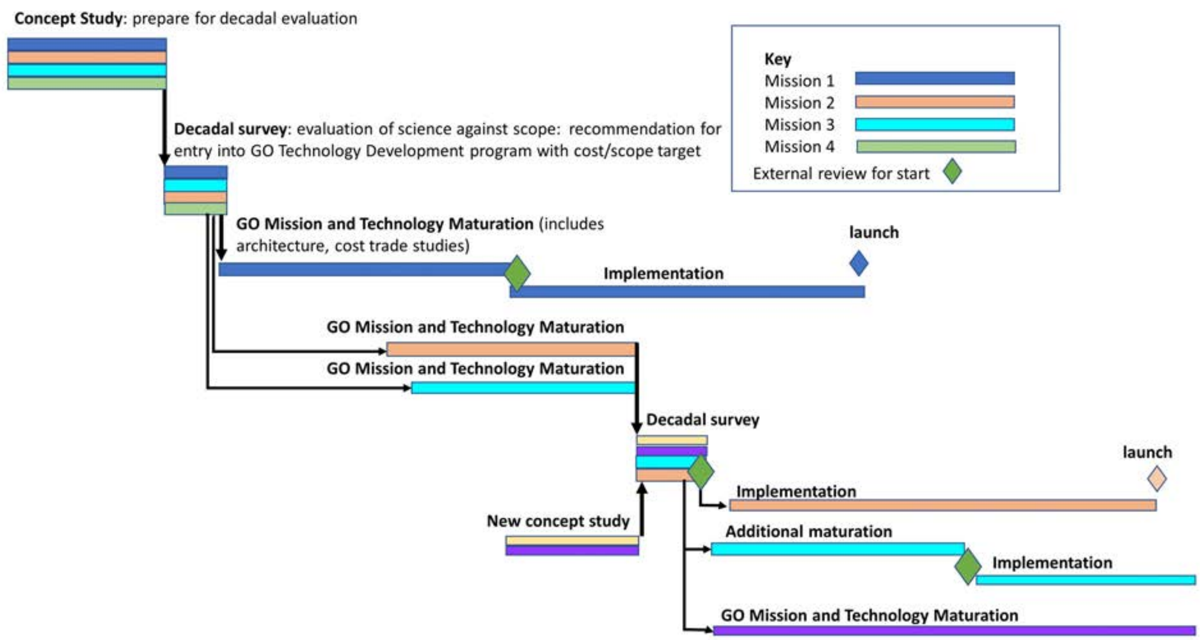


FIGURE 7.2 Flow diagram showing the concept for maturation, recommendation and implementation of NASA Large Strategic Missions. This does not represent the actual recommendations in this report, instead it represents how the program would be structured in general. If implemented, this survey would be the first to adopt this process by recommending the first entrant into the GO Mission and Technology Maturation Program. SOURCE: Fiona Harrison.

> **Recommendation: The NASA Astrophysics Division should establish a Great Observatories Mission and Technology Maturation Program, the purpose of which is to co-develop the science, mission architecture, and technologies for NASA large strategic missions identified as high priority by decadal surveys.**

For recommended missions, decadal surveys would provide guidance, where possible, on the mission cost target, and scientific scope. Key elements of the program, summarized in Figure 7.3, include:

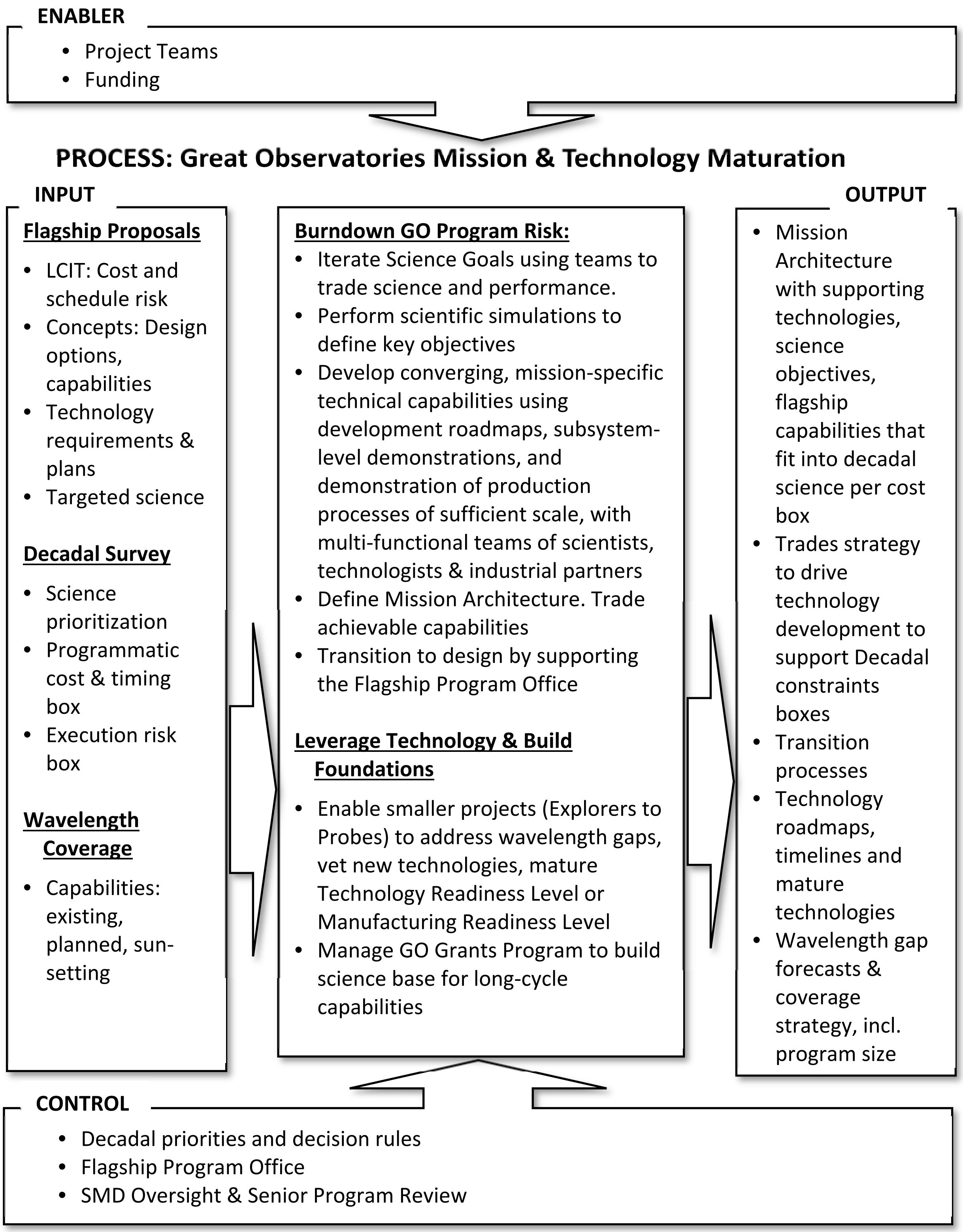


FIGURE 7.3 Flow diagram showing the key functions and scope of the Great Observatories (GO) Mission and Technology Maturation Program. SOURCE: Sigur/Harrison.

- Mitigating the program risk associated with implementation of new large observatories by iterating and defining the science using broad community engagement, including using teams and open grants competitions, to evaluate science and performance trades and undertake scientific simulations for key objectives.
- Developing the mission architecture best suited to accommodate programmatic constraints, while also supporting optimized science objectives. This would be accomplished using vetted and reviewed development roadmaps, subsystem-level demonstrations, and prototyping of production and manufacturing processes at a sufficient scale, with multi-functional teams of scientists, technologists and industrial partners.
- Transitioning collaboratively from The Great Observatories Maturation Program Office which oversees feasibility studies and technology demonstrations, to design and production, as recommended by a decadal survey or mid-term decadal assessment. At this review it could be decided not to advance a mission due to cost or other factors.

For this decade, the following missions are advanced for inclusion in this program:

The survey's top priority for this program is an **IR/O/UV telescope optimized for observing habitable exoplanets and general astrophysics**. As described in more detail in Section 7.5.2, the mission is recommended for implementation later in the decade, but only after the successful completion of the associated Great Observatories Mission and Technology Maturation program. Based on TRACE analysis and program panel input, the Survey estimates that 6 years will be required for this maturation, starting as soon as possible, before the mission is ready to be considered for implementation. The estimated cost of this mission and technology maturation program is ~$800 million, based on the cost and schedule analyses from the TRACE for the LUVOIR-B technology maturation program. These costs are carried within the Great Observatory Mission and Technology Maturation Program for approximately 6 years, at which point any residual technology development and the associated costs are transferred to the IR/O/UV mission development line. If this schedule and funding level can be achieved, by late decade it will be possible to assess the mission design, scientific reach, technology readiness at both the component and system level, feasibility of manufacturing processes, and cost for consistency with the survey's recommendation and NASA's budget guidance prior to transitioning to formulation and implementation.

After an IR/O/UV exoplanet and astrophysics mission enters formulation, the survey assigns equal priority to commencing mission maturation and technology development for a **far-IR spectroscopy and imaging strategic mission**, and a **high spatial and spectral resolution X-ray strategic mission**. The survey committee believes an appropriate cost target for implementing these missions is $3 billion–$5 billion (FY2020) each. The motivation for this cost target is to balance the scientific scope with timeliness. Unlike the IR/O/UV mission, which has a target aperture based on a requisite number of Earth-like exoplanet detections, which in large part drives the cost, the Far-IR and X-ray missions are more easily scalable. In the survey committee's judgement, missions with simplified design and selected instrument capabilities relative to Origins and Lynx can make substantial progress in addressing the priority science themes. A lower cost target relative to the TRACE estimates in Table 7.4 will enable these missions to advance more rapidly to implementation and realization, so that they can potentially have some overlap in operational lifetime. The survey committee views this as more important than achieving the full scope of Origins or Lynx. Finally, the scientific focus on the co-evolution of black holes and galaxies suggested by the EOS-2 panel is not necessarily the correct one. Rather, the mission maturation program would include trade studies to determine the scientific foci that are consistent with the broad set of the survey's identified science priorities as well as the suggested cost target.

An appropriate funding level for each of these programs is $40 million a year beginning in the second half of the decade. This is based on the level of mission and technology development funds

described in the concept studies (~$600 million for Lynx, ~$350 million for Origins), and the expectation that a significant fraction of at least 20-30 percent of this will be required to mature the missions and technologies to the appropriate level to enable a recommendation for mission implementation in the 2030 decade. If available budget levels require a choice to be made about which mission enters the program first, the survey committee suggests that the mid-decadal review evaluate the international scientific landscape, outcome of the probe selection, and that this review provide advice on which should commence maturation first.

7.5.2 Frontier Projects: A Future Large IR/Optical/UV Telescope Optimized for Observing Habitable Exoplanets and General Astrophysics

Exploring terrestrial planets outside our Solar System through direct imaging and spectroscopy will advance one of humanity's greatest quests - the search for habitable environments and life outside of the Solar System. This transformative goal is at the scientific forefront, connecting astronomy, astrobiology, planetary and Earth Science, and is one that captures the imagination of all humankind. Searching for signatures of life on potentially habitable planets around stars like the Sun, and exploring entire solar systems beyond our own (Figure 7.4) can only be achieved with a large aperture space telescope optimized for this purpose.

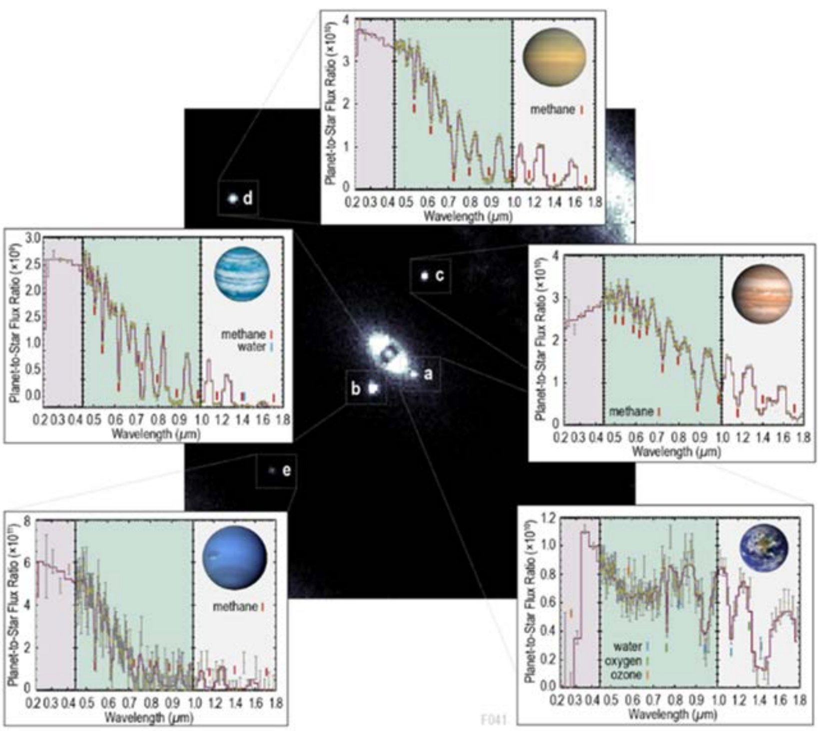


FIGURE 7.4 Simulated space-telescope image of a complete planetary system including a life-bearing Earth-like planet. Simulated planets are shown clockwise from bottom left: e) close-in Neptune, b) sub-Neptune, d) Saturn, c) Jupiter, and a) exo-Earth. The system also contains inner and outer dust belts. Spectroscopic studies would allow the mission to distinguish between the planets and explore the complete evolutionary history of the system. SOURCE: NASA, HabEx report. The Habitable Exoplanet Observatory Study Team.

A mission reaching this goal is challenging technically, scientifically and programmatically, yet, given rapid progress over the past 10 years, commencing its design and development is now possible. In the last decade the uncertainty in the number of Earth-sized potentially habitable planets has been reduced

by Kepler and other missions, and it is now known that such planets are common. Improved understanding of the complexities of planetary atmospheres lets us identify the spectroscopic measurements needed to assess the signatures of life. The technological feasibility of blocking starlight to see planets ten billion times fainter than their host stars has been demonstrated in laboratory testbeds, and although not at the level required by the IR/O/UV observatory envisioned here, several key capabilities will be tested by the Roman Space Telescope. We are on the threshold of a transformational leap in capability that will enable not just discovery but also exploration of planets beyond our Solar System. The key pathway to finding new, habitable worlds leads directly through this IR/O/UV space observatory.

The same large aperture telescope that can identify Earth analogs would be equally transformative for general astrophysics. Its broad wavelength coverage, extending from the ultraviolet through the visible into the near-infrared, would inherit the scientific power of the Hubble Space Telescope, but with a light collecting area several times larger, 2-3 times sharper image quality, and instruments and detectors significantly more sensitive, providing 1-2 order-of-magnitude leaps in sensitivity and performance over HST. This telescope will be capable of achieving breakthrough discoveries across nearly all of astrophysics. Prime examples include ultraviolet and visible spectroscopy of the circumgalactic halos and the intergalactic medium and of mass flows within and out of galaxies to reveal the workings of cosmic ecosystems in detail and depth for the first time; high-resolution observations of supermassive black holes and their host galaxies locally and over cosmic time; and the construction of stellar fossil histories of the galaxies in the neighborhood of the Milky Way. The nature and effects of dark matter can be addressed by measuring the joint 3-dimensional kinematic and dark matter density profiles of dwarf galaxies. These examples all constitute major components on the Dynamic Universe and Unveiling the Drivers of Galaxy Growth priority science, and they only represent the tip of the iceberg of the impact such a telescope would have. This observatory will become one of the most scientifically versatile astronomical telescopes ever flown, and its observations will directly address two-thirds of the 24 key science questions identified in Chapter 2 and will contribute towards answering many of the others.

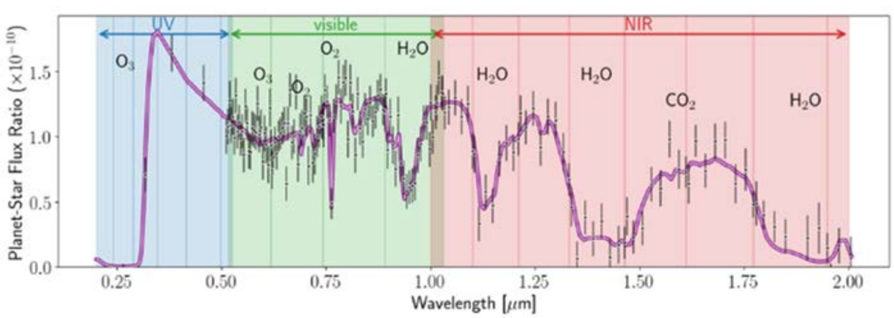


FIGURE 7.5 Simulated spectrum of an Earth-twin planet observed from the UV to near-IR by a space coronagraph. Spectral features from oxygen, water, ozone and CO_2 show the presence of a biosphere. SOURCE: NASA LUVOIR report. J. Lustig-Yaeger (University of Washington).

The mission the survey puts forward will combine a large, stable telescope with an advanced coronagraph intended to block the light of bright stars. It will be capable of surveying a hundred or more nearby Sun-like stars to discover their planetary systems and determine their orbits and basic properties. Then for the most exciting ~25 planets, astronomers will use spectroscopy at ultraviolet, visible, and near-infrared wavelengths to identify multiple atmospheric components that could serve as biomarkers (see Figure 7.5). It will also have high-resolution imaging and multi-object spectroscopic capability (particularly at ultraviolet wavelengths) to address a broad range of astrophysical science selected through guest investigator programs.

After considering the analysis from the EOS-1 panel regarding technology readiness, cost, and science capability, and weighing the need for program balance and timeliness, the survey committee

concludes that a high-contrast direct imaging mission with a target off-axis aperture of approximately 6 meters provides an appropriate balance between scale and feasibility. Such a mission would yield a robust sample of ~25 atmospheric spectra of potentially habitable exoplanets, and it could launch by the first half of the 2040 decade. A sample this size provides robustness against the uncertainties in the occurrence rate of Earth-sized worlds, and against the vagaries associated with the particular systems near Earth. Analysis by the EOS-1 panel finds that, given the budget requirements and realistically achievable yearly funding levels, an 8 m aperture telescope of the scale of LUVOIR-B would be unlikely to launch before the late 2040's or early 2050's. On the other hand, a smaller telescope such as the HabEx 4H design may fall short of providing a robust exoplanet census, and was judged by EOS-1 not to advance general astrophysics capabilities sufficiently to justify the high cost and long timescale for completion.

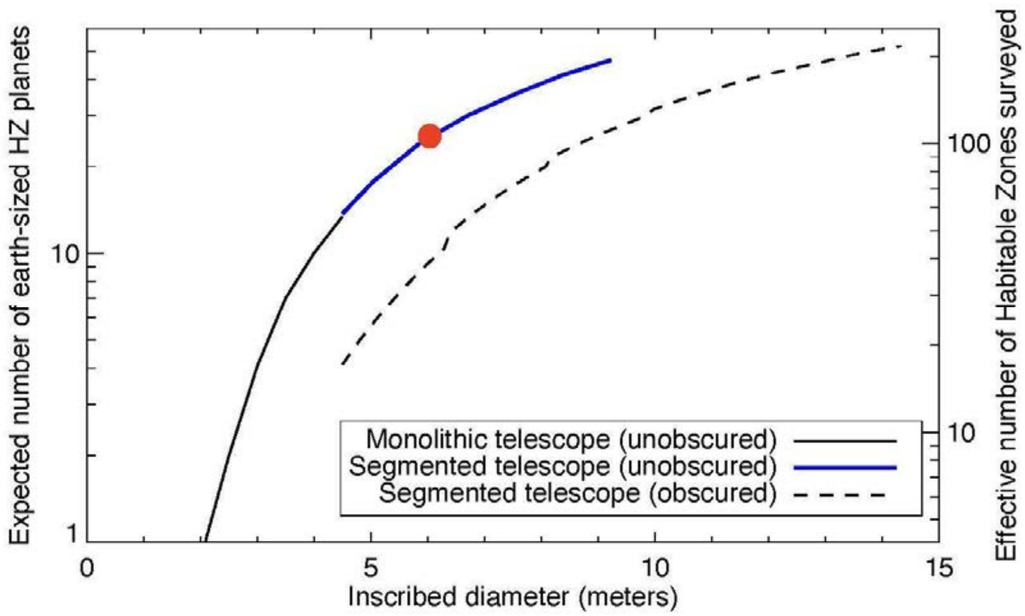


FIGURE 7.6 Potentially habitable exoplanet yield vs telescope diameter for different telescope architectures. Right axis shows the number of habitable zones surveyed (weighted by completeness); left axis shows the expected number of planets discovered assuming the occurrence rate of rocky planets in the optimistic habitable zones of different stars, eta_earth=0.24 (Bryson et al. 2021). The red dot shows the expected yield for the target 6-m inscribed diameter. NOTE: Habitable zone is defined as 0.95-1.67 AU for planets of 0.8-1.4 Earth radii. SOURCE: Adapted from C. Stark (Space Telescope Science Institute), D. Mawet (California Institute of Technology), and B. Macintosh (Stanford University).

> **Conclusion:** A high-contrast direct imaging mission with a target off-axis inscribed diameter of approximately 6 meters provides an appropriate balance between scale and feasibility. Such a mission will provide a robust sample of ~25 atmospheric spectra of potentially habitable exoplanets, will be a transformative observatory for general astrophysics, and given optimal budget profiles it could launch by the first half of the 2040 decade.

Realizing this mission requires significant technology development and maturation of the design and implementation. The best path forward is to have NASA immediately commence aggressive technology development aimed at achieving the goal described above as part of the Great Observatories Mission and Technology Maturation Program. This program would consider and optimize configurations targeted at performance consistent with the target 6-m off-axis aperture as indicated in Figure 7.6. These studies would combine scientific and technical ideas and talent from the entire community to develop a

single mission architecture and associated technology roadmap. Broad participation of astronomers, exoplanet scientists, and Earth scientists would be beneficial in refining the exoplanet observation strategy. As noted above, even if begun immediately, the technology development phase is likely to take 6 years, and possibly more, and could require an investment of $800 million prior to the Phase A start of the resulting mission.

Given current uncertainties in how key technologies will advance, and what the budget climate will be for implementing an ambitious strategic mission, the survey recommends that the trade between launch date and mission scale be reviewed again after the crucial technologies have reached maturity (technology readiness level [TRL] 5-6), and the mission implementation, descopes and trades are better understood. At this time key scientific inputs, such as the frequency of Earth-sized planets in habitable zones of stars, will also be better constrained by ongoing observational work, and the mission capabilities, cost, and schedule will be significantly refined. Discussions about key international partnerships on this ambitious program can also be established at this time. Prior to commencing formulation, the proposed implementation would be subject to an independent review which could assess plans in light of budgetary needs and realities, and subsequently NASA could move forward with implementing a mission that accomplishes transformative exoplanet and general astrophysics goals while being realizable by the first half of the 2040 decade.

Recommendation: After a successful mission and technology maturation program, NASA should embark on a program to realize a mission to search for biosignatures from a robust number of about ~25 habitable zone planets and to be a transformative facility for general astrophysics. If mission and technology maturation are successful, as determined by an independent review, implementation should start in the latter part of the decade, with a target launch in the first half of the 2040s.

This is an ambitious strategic mission, and while not at the cost scale of LUVOIR, it will still require an investment comparable to Hubble or JWST. To assess the budget scale and profile requirements for the recommended direct imaging mission, the survey committee performed an analysis assuming the cost profile and schedule from the LUVOIR-B TRACE analysis, normalized to a total integrated cost equivalent to JWST inflated to current year dollars.[11] The survey committee believes this is a conservative assumption: the JWST telescope incorporates a 6.5 m segmented primary mirror, and as it operates in the mid-infrared it has many tight thermal requirements. A ~6 m aperture high contrast imaging mission would have the added complexity of extreme starlight suppression, but would operate at and could be tested at room temperature. Many factors, including insufficient technology development prior to Phase A start, sub-optimal funding profiles, and management and execution challenges increased the overall cost of JWST relative to what should be achievable. Here the survey is recommending an aggressive early technology phase so that development can proceed optimally after a mission start. The result of the analysis is that a total cost of $11 billion (FY2020), with a real-year funding profile shown in Figure 7.10, is likely to bound the mission cost.

The survey committee believes that the scientific goals of this mission, when achieved, have the potential to profoundly change the way that we as human beings view our place in the universe. With sufficient ambition, we are poised scientifically and technically to make this transformational step. This endeavor represents a quest that is on the technical forefront, and is of an ambitious scale that only NASA can undertake, and it is one where the United States is uniquely situated to lead the world.

[11] LUVOIR-B was chosen as a basis for the estimate because it was more scalable with its segmented mirror, internal coronagraph, and single launch vehicle.

7.5.3 Sustaining Activities

The two recommendations below are aimed at capitalizing on the upcoming Roman, Rubin, Athena and LISA observatories, and balancing scientific progress among the survey's priorities, thereby addressing the extraordinary richness of 21st century astrophysics.

7.5.3.1 Time Domain Astrophysics Program

Exploring the cosmos in the multi-messenger and time domains is a key scientific priority for the coming decade. The ability to probe time-variable, explosive and transient phenomena has been propelled by large format detectors, by dramatic computational advances, and in the last decade by the advent of entirely new means of discovering transient phenomena through gravitational waves and high energy neutrinos. As a natural result of their large fields of regard and cadenced observations, time domain observations are a central element of the top ground and space-based projects supported by Astro2010; the Vera Rubin Observatory (referred to as the Legacy Survey of Space and Time [LSST] in 2010); and the Nancy Grace Roman Space Telescope (WFIRST in 2010). Time-domain studies offer tremendous new discovery space for both observatories, as well as for many other U.S. space and ground-based facilities (e.g., the Transiting Exoplanet Survey Satellite [TESS], Zwicky Transient Facility [ZTF], Deep Synoptic Array 110 [DSA-110], the Laser Interferometer Gravitational-wave Observatory [LIGO], and IceCube).

Many of the exciting opportunities in multi-messenger astrophysics rely on observations of astrophysical transients discovered via gravitational waves, and neutrinos, but requiring electromagnetic observations across the spectrum to identify and probe the fundamental physical nature of the phenomena. The LIGO/Virgo-discovered binary neutron star merger, GW170817 is a prime example (Chapter 2, Box 2.2). Looking to the future, the LISA mission will open enormous discovery space for probing larger mass black holes, as well as white dwarf binaries, where electromagnetic observations will be equally essential.

While ground-based measurements by observatories large and small are essential, several key capabilities that must be sustained to enable time-domain and multi-messenger astrophysics can only be realized in space. The most important of these are wide-field gamma-ray and X-ray monitoring, and rapid and flexible imaging and spectroscopic follow-up in the X-ray, ultraviolet (UV), and far-infrared (far-IR). In addition, space platforms can be designed to access much of the sky at any given time, essential for the study of short-lived transients or rapidly variable sources. Space missions can also observe near-continuously compared to ground-based telescopes.

As discussed in the report of the EOS-2 panel, many of the necessary observational capabilities can be realized on Explorer-scale platforms (Missions of Opportunity (MoO), Small Explorers (SMEX) and Medium-class Explorers (MIDEX), while others could require larger efforts, but still less than half the scope of a Probe Mission. In addition to the threats of lost capabilities resulting from the aging of Swift and Fermi, there are also potential new international opportunities to meet the scientific needs, such as the Space Variable Objects Monitor (SVOM). Contribution of instruments to international efforts is another possibility for achieving some elements of the program. The specific needs to sustain and enhance the optimum suite of space capabilities will change over the upcoming decade, and it is likely that these capabilities will be most effectively achieved by a complement of missions on different scales, including contributions to international efforts.

One effective mechanism for achieving the above goal would be for NASA to appoint a standing planning committee in time-domain astrophysics. This committee could, over the decade, periodically review and advise NASA about what the critical needs are to maintain and expand a vibrant and effective system of time-domain and transient follow-up observatories, evaluated in the international context, and considering what can effectively be done by ground-based facilities. The committee's considerations would be guided by the key scientific questions identified by the survey. NASA could then respond to

these needs either through targeted calls that are part of the Astrophysics Research and Analysis Program (APRA) or the Explorer program, or through dedicated announcements of opportunity.

Conclusion: A standing planning committee or advisory structure could provide tactical advice to NASA on impending needs and priority capabilities for time domain and multi-messenger follow-up, evaluating these needs in the international landscape.

Recommendation: NASA should establish a time-domain program to realize and sustain the necessary suite of space-based electromagnetic capabilities required to study transient and time-variable phenomena, and to follow-up multi-messenger events. This program should support the targeted development and launch of competed Explorer-scale or somewhat larger missions and missions of opportunity.

The estimated cost of this program would range from $500 million–$800 million over the decade. This lower range would support competed missions of opportunity, SMEX and MIDEX scale missions. As described in Section 6.2.1.1.3, the survey notes this funding is intended to be added above the current funding level of the Explorer program, so as not to negatively impact the rate of selections through entirely open, non-targeted calls. These expenditures would take place throughout the decade.

7.5.3.2 Astrophysics Probe Mission Program

Advancing the survey's science program depends on the existence of a panchromatic suite of capabilities. Given the long development timescales for large strategic missions, establishing a 'probe' class line with mission costs of ~$1.5 billion and launches every decade, will address the need for broad wavelength coverage and scientific balance. Through advances in technology, combined with focused science, missions at this scale can achieve more than an order of magnitude leap in capability, and address scientific areas of high priority. This is supported by the large number of white papers for missions in this category evaluated by EOS-1 and EOS-2. Both panels concluded that probe-scale missions offer exceptional scientific opportunities. Missions at this scale would also address the significant gap in cost, capability, and development time scales between Explorers and strategic missions.

A mission cost-cap of $1.5 billion affords an appropriate balance between capability and launch cadence. For many of the projects considered by the program panels, a cost cap of $1 billion was constraining. This conclusion is supported by NASA's Probe Cost Assessment Team report, which found that only one of the nine NASA-supported probe concept studies is likely to be achievable within a $1 billion cap.[12] At the other extreme, pushing the cap to $2 billion would decrease the launch rate to less than one per decade, given other demands on the program. In contrast to entirely open competitions (such as in the Explorer program), probe class mission calls that target areas where there are strategic scientific gaps in decadal programs will advance survey priorities in a more optimal way. A good model for achieving this is the approach taken by NASA's Planetary Science New Frontiers Mission line.

Conclusion: The large gap in cost and capability between Explorer class missions and large strategic missions is a significant impediment to achieving the broad set of Astro2020 decadal scientific priorities. Institution of a Probe class line of missions, with an individual mission cost cap of ~$1.5 billion, selected from priority areas identified by decadal surveys, would broaden NASA's astrophysics program in a way that better addresses the richness of 21st century astrophysics.

[12] NASA, "PCAT Final Report," 2019, Independent Probe Cost Assessment Team (PCAT), https://science.nasa.gov/science-red/s3fs-public/atoms/files/PCAT_FinalRpt-2019-Nov8.pdf.

Recommendation: NASA Astrophysics Division should implement a line of probe missions with a mission cost cap of ~$1.5 billion fiscal year 2020 and a targeted launch rate of approximately one per decade. These missions should be competed, with solicitations calling for concepts in priority areas identified by decadal surveys.

The survey identified two priority areas as the most compelling for probe missions this decade. These are highlighted, because they address crucial scientific gaps, and wavelength ranges and/or observational capabilities important to the survey's scientific objectives, but where no mission is currently planned either nationally or internationally. In calling out these areas, the survey is not endorsing any of the specific mission concepts from submitted white papers, or any of the NASA probe studies. Rather, the probe implementation will require NASA to support the development and study of concepts that fit within the cost cap, most likely through a concept study/down select process.

The two areas for the first probe competition are listed below. Allowing both areas to compete for the first opportunity will enable a robust number of mission concepts, providing a high likelihood that a compelling mission is selected and developed within the cost cap of the program. A third area is listed where investment in technology development this decade would prepare for a subsequent probe call early in the 2030's.

7.5.3.3 A Far Infrared Imaging or Spectroscopy Mission

A far-IR imaging or spectroscopy probe mission would address scientific objectives central to Astro2020, and would fill an important gap in world-wide capabilities. Since the EOS-2 report was completed, ESA made the decision to remove the joint ESA/Japan Aerospace Exploration Agency (JAXA) Space Infrared Telescope for Cosmology and Astrophysics (SPICA) far-IR mission from consideration for its M5 slot. SPICA was identified as a priority for NASA participation by Astro2010, would have flown a powerful set of spectrometers covering the 12 – 230 μm range, as well as a mid-infrared imager. SPICA was positioned to make significant progress in a number of the science areas highlighted by this survey. Recent improvements in far-IR technology mean that major scientific advances can be made compared to the Herschel Space Observatory.

The EOS-2 panel considered the landscape for a future far-IR mission prior to ESA discontinuing its consideration of SPICA. The survey committee believes that considering this change in landscape there are many unique opportunities for a properly scoped far-IR probe to advance high priority science, and a probe scale mission is an extremely timely and compelling opportunity to do so. These scientific areas include tracing the astrochemical signatures of planet formation (within and outside of our own Solar System), measuring the formation and buildup of galaxies, heavy elements, and interstellar dust from the first galaxies to today, and probing the co-evolution of galaxies and their supermassive black holes across cosmic time. These goals are all central to the broader scientific themes of the survey. The ultimate scientific focus of the far-IR probe will depend on the outcome of the competitive selection.

7.5.3.4 An X-ray Probe to Complement ESA's Athena Observatory

Observations of the universe in the X-ray band probe many energetic phenomena and processes not accessible through other wavebands. The ESA L-class *Athena* X-ray mission has many of the capabilities for a next-generation X-ray observatory that were recommended by Astro2010. With its very large aperture, moderate spatial (5") resolution imager, and non-dispersive spectrometer (ΔE of 2.5 eV) operating from 0.1 – 8 keV, *Athena* will be a major step forward. However, Athena lacks the high spectral resolution (R ~ 7500), broad bandpass, and spatial (<1") resolution needed to address multiple Astro2020 key science questions. These include understanding the invisible drivers of galaxy formation and evolution through observations of effects of feedback from accretion of gas onto supermassive black

holes, searching for the first seed black holes at high redshifts (z~10), and characterizing the activity of stars and studying their evolution.

Because of the unique science that can be addressed with a mission complementary to Athena, a targeted X-ray probe is one of the priorities for a probe mission competition. While a probe would not fulfill all objectives of a large strategic mission such as Lynx, a probe mission could enhance Athena, and also address important capability gaps. Which of the scientific objectives mentioned above can be achieved by a probe scale mission is unclear at this time. For a mission properly selected and scoped, there are multiple potential science opportunities that are both unique and timely. Depending on the scientific focus, it is not necessary for the X-ray probe to be operational simultaneously with Athena, rather the survey envisions strong complementarity of the science focus.

7.5.3.5 An Early Universe Cosmology and Fundamental Physics Probe

As detailed in the report of the Panel on Cosmology, studies of the cosmic microwave background continue to provide data that address profound and fundamental questions about the universe on the largest scales and during its earliest moments. As noted by the EOS-2 panel report, “space observations will unquestionably be needed for the best foreground separation and the lowest systematic errors on all angular scales, and especially on angular scales of greater than about ten degrees.” With investment in technologies this decade, combined with ground-measurements, cosmic microwave background (CMB) probe mission could potentially be a compelling candidate for the future probe call in the 2030’s, complementing the survey’s ground-based CMB-S4 recommendation.

7.5.4 Prioritization of NASA Sustaining and Frontier Activities

The survey’s top priority for medium and large programs is for NASA to complete the major astrophysics facilities and missions currently in development, including its commitments for participation in major ESA missions, and to maintain the Explorer program at the current healthy rate (see Table 7.1 for a list of these activities).For new missions and programs, the survey does not prioritize projects between sustaining activities and advancing large strategic missions, as both achieve equally important goals for the program. This parallels previous surveys that have not prioritized programs in the large category relative to medium or small scale activities. The rationale for this is the overriding need for the balance of mission and program scales required for the success of the astronomy and astrophysics enterprise. In the sustaining program category the survey prioritizes the time-domain program over the probe line, due to the urgent need to maximize return from major U.S. investments in LIGO and Rubin Observatory on the ground, and Roman in space. For large strategic missions, the highest priority is for NASA to rapidly establish the Great Observatories Mission and Technology Maturation program, with the most important element in that category being to commence maturation of the large IR/O/UV mission.

The largest budgetary increase associated with the recommended program arises in the latter half of the decade, assuming that the large IR/O/UV mission is technically ready and sufficiently mature to commence detailed design and implementation (See Figure 7.10). If, for budget or technical reasons, this must be delayed, it is still important for NASA to proceed with the mission and technology maturation programs for a Far IR observatory and a high-resolution X-ray observatory. This will help to minimize the time between the ultimate realization of the IR/O/UV mission and the subsequent large strategic mission.

7.6 ROADMAP FOR GROUND: NEW ACTIVITIES

The sections below describe the new ground-based activities that the survey recommends that NSF and DOE undertake in the next decade, divided into three categories: sustaining activities that broaden science and the timescales on which new capabilities emerge; enabling activities that advance future MREFC-scale facilities; and the frontier observatories that are ready for implementation in the coming decade, and that will ensure the U.S. community continues to advance the scientific forefronts.

7.6.1 Frontier Facilities and Observatories

Large observatories provide the transformative capabilities that achieve breakthrough discoveries, and advance a broad range of scientific objectives. The survey recommends phased NSF investment in three large programs this decade. The highest priority is participation in the United States. Extremely Large Telescope (ELT) Program, because of its transformative nature, and because of the urgency of this investment to the projects' success. Next, equal priority is placed on developing and implementing the CMB-S4 observatory together with the DOE, and on beginning design, cost studies and prototyping for the next-generation Very Large Array radio telescope. Finally, if these studies are successful, and if budgets allow, the survey recommends commencing construction of the next generation Very Large Array (ngVLA) toward the end of the decade. While the IceCube Gen-2 neutrino observatory is not funded out of AST, an assessment is provided of its relevance to the science recommended by this survey.

7.6.1.1 The U.S. Extremely Large Telescope Program

The U.S. ELT Program as proposed to the survey is made up of three elements: the Giant Magellan Telescope (GMT), the Thirty Meter Telescope (TMT), and NSF's National Optical-Infrared Astronomy Research Laboratory (NOIRLab) (See Figure 7.7). The primary mirror of the GMT has a total diameter of 24.5 meters and the telescope has a 25 arcmin field-of-view (FOV). The GMT will be located at the Las Campanas Observatory in Chile. The majority of the GMT partners are U.S. institutions, with international partners in Australia, Brazil, and Korea. The TMT primary mirror has a diameter of 30 meters and the telescope has a 20 arcmin FOV. The TMT will either be sited on Maunakea in Hawaii, or at Roque de los Muchachos Observatory on La Palma in the Canary Islands. The majority of the TMT partners are international, with the participation of institutions in the United States, Canada, China, India, and Japan. For comparison, the European Southern Observatory (ESO) is building the ESO ELT on Cerro Armazones in Chile with a 39.3 m diameter and a 10 arcminute FOV, with first light expected in 2028. Both the TMT and the GMT are well into development; both projects have mature designs and have commenced fabrication of key elements, although challenges remain. They are expected to commence operations in the mid 2030's, contingent on a U.S. funding commitment.

The scientific potential of the ELTs is vast. The combination of large collecting area (4-9 times that of a 10m Keck telescope) and diffraction-limited imaging (0.01-0.02" FWHM with adaptive optics in the near-IR) provides observational capabilities unmatched in space or the ground, and opens an enormous discovery space for new observations and discoveries not yet anticipated. A resolution of 0.01" (12 times that of the Hubble telescope at similar wavelengths) projects to a linear scale of 25 km at Jupiter, 1 AU for a protoplanetary disk at distance 100 pc, 0.8 pc at the distance of the Virgo Cluster, and 60 pc for galaxies at redshift $z=2.5$ (comparable to the scales resolved by ground-based telescopes with natural seeing at Virgo). For unresolved sources the sensitivity of these telescopes scales as their diameter to the fourth power, a gain of 36–81 times over current 10 m telescopes. The large collecting areas of the telescopes also makes them powerful spectroscopic machines, especially for high-resolution spectroscopy where measurements are often limited by detector noise.

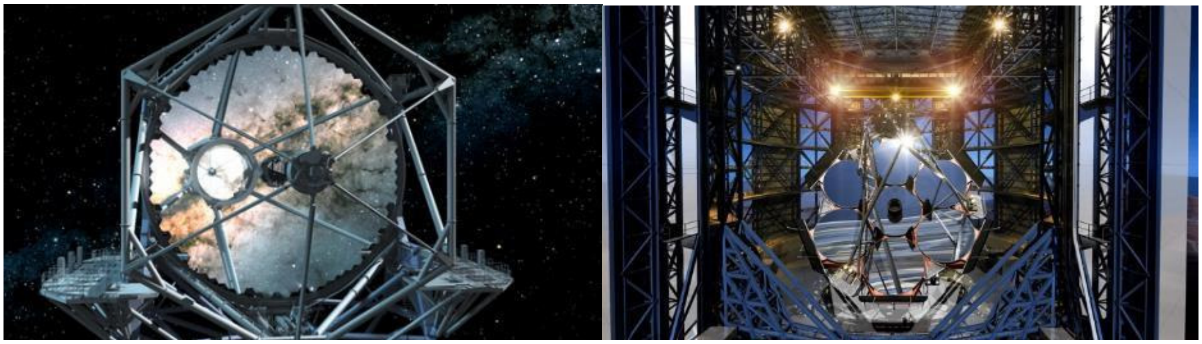


FIGURE 7.7 The Thirty Meter Telescope mirror (left) and the Giant Magellan Telescope mirror (right). Both ground-based projects are at an advanced stage of development and have commenced construction of their primary optics. SOURCE: *Left:* https://www.tmt.org/page/uselt. Courtesy of the TMT International Observatory. *Right:* Courtesy of the Giant Magellan Telescope – GMTO Corporation.

This powerful combination of capabilities can be brought to bear on nearly all of the important science questions laid out by this decadal survey, across all three of our key science themes.[13] They will be able to detect, image, and characterize temperate rocky planets around low-mass stars, measure their atmospheric compositions including searches for oxygen, image protoplanetary disks, and through precision radial velocity measurements measure the masses of the planets, vital information only possible with the ELTs. Fundamental physics will be probed through a variety of pathways, including measurements of stars orbiting the Milky Way's central black hole SgrA*, to perform tests of relativity and gravity. Measurements of the cosmic expansion rate using different methods (variable stars, gravitational lensing, merging neutron star "standard sirens") will test for the reality of the current Hubble tension and reveal whether the current ΛCDM cosmological model fully describes the expansion.

Measurements of the faint spectra of gamma-ray bursts and supernovae beyond redshifts z=10 will probe both the physics of stellar explosions at early cosmic times and probe this epoch of reionization itself. The impacts of the ELTs for revealing the workings of the Cosmic Ecosystem promise to be especially powerful. These telescopes alone will have the sensitivity to make spectroscopic measurements of the faintest galaxies, stellar explosions, and black holes detected by JWST; the result of these studies will be a record over cosmic time of the buildup of matter, stars, heavy elements, and the assembly of the galaxies themselves from hundreds of thousands of years after the Big Bang to the present. Likewise, these telescopes will have the unique ability to trace the chemical and dynamical buildup of the Milky Way and nearby galaxies out to the Virgo cluster, through deep high-resolution imaging and spectroscopy of their oldest stars. Many of these unique capabilities complement those of our top-ranked space project, the IR/O/UV space telescope, extending the powerful synergies between the ground-based 6-10 m telescopes and HST over the past 30 years (and soon with JWST and the Roman telescope). As demonstrated by the 16 6.5-12m OIR telescopes currently in operation around the world (not counting the Rubin Observatory or others under construction), the versatility of these instruments and the large range of top-priority scientific applications will more than fully occupy even three ELTs for decades.

> **Conclusion:** Because of their transformative scientific potential, as well as readiness, the success of at least one U.S. ELT is a critical priority for investment for ground-based astronomy in the coming decade.

Although the U.S. astronomy community would benefit enormously by an NSF investment in even one of the TMT or GMT, there is considerable benefit to pursuing a coherent two-telescope U.S. ELT Program that would combine capabilities of both A two-telescope U.S. ELT program would offer

[13] See Table 1 in the report of the Panel on Optical and Infrared Observations from the Ground.

full-sky coverage, important for leveraging the current U.S. multi-billion dollar bi-hemispheric system of ground-based OIR and radio astronomical facilities (JVLA and the future ngVLA in the north, ALMA and the Vera Rubin Observatory in the south) and assure observations of rare objects (e.g., nearby habitable exoplanets, rare classes of transient events) regardless of where they lie in the sky. Complementary instrumentation on the two telescopes developed in a coherent manner in partnership with NOIRLab would significantly increase the scientific reach of the overall U.S. ELT program. Investing in two telescopes would also maximize the total number of nights of public-access observing time—potentially as much as 200 nights per year—and far more than remain available for NSF partnership on either of the observatories alone.

The enormous scientific potential of the ELTs has also been recognized overseas. Several international organizations are partners in the GMT and TMT project, and in 2008 a European Astronet decadal study identified an ELT as one of its top priorities (along with a Square Kilometer Array radio telescope project). ESO now is constructing a 39 m ELT in Chile, with planned commissioning later in this decade. NSF participation in a U.S. ELT program will position the U.S. community to take full advantage of the promise of these facilities. Although smaller in aperture the TMT and GMT offer a number of unique capabilities, including fields of view 4-6 times larger than the ESO ELT (facilitating multi-object spectroscopy), and high-resolution first-generation spectrometers capable of carrying out groundbreaking observations of exoplanets, ancient stars, and the circumgalactic and intergalactic media, key elements of the Habitable Worlds and Galaxy Growth priority areas. These capabilities are regarded less as competitive advantages than as powerful synergies between complementary facilities which will hasten the advancement of the science frontier objectives highlighted in this survey.

As proposed to this survey, the U.S. ELT program would be comprised as a collaboration between the GMT and TMT projects with the NSF NOIRLab. NOIRLab would provide proposer and user support, public data products and archiving, broaden participation in U.S. ELT science, foster research inclusivity, and engage and represent the whole U.S. community in the U.S. ELT governance and scientific planning. NSF partnership would leverage major investments by universities and foundations ($1.5 billion), and international partners ($1.2 billion), and assure that the fruits of these revolutionary facilities are shared by the largest possible community of researchers and students in the United States.

The Panel on Optical and Infrared Observations from the Ground (OIR) assessed the programmatic and technical risks and cost of both the GMT and TMT separately, and both underwent an independent TRACE analysis. The TRACE construction cost estimates of $2.4 billion and $3.1 billion for GMT and TMT respectively are within 20 percent of the project cost estimates ($2 billion and $2.65 billion), which is within the uncertainties. While there are technical challenges for both projects, solutions appear to be in-hand. TMT has the added risk that the site has not yet been selected, adding cost and schedule uncertainty. However, the biggest risk for both projects is the large gap between commitments in-hand from the partners, and what is required to complete the projects, even with a significant federal investment by NSF of $0.8 billion per project. This programmatic risk is significant, and the TRACE analysis gave both projects a medium-high programmatic risk rating.

The scientific potential of the ELTs is so compelling, and the science so broad, that ideally community access would be at least 25 percent on each of the ELTs (as proposed to the survey). If, however, programmatic or financial challenges preclude the viability of one of the projects, the survey recommends that NSF invest in at least one ELT, with a share of the time proportional to the fractional federal investment in constructions and operations.

Recommendation: The National Science Foundation (NSF) should achieve a federal investment in at least one and ideally both of the two extremely large telescope projects—the Giant Magellan Telescope and the Thirty Meter Telescope, with a target level of at least 25 percent of the time on each telescope. If only one project proves to be viable, NSF should aim to achieve a larger fraction of the time, in proportion to its share of the costs and up to a maximum of 50 percent.

7.6.1.2 Criteria and Decision Rules for Investment in the U.S. ELTs

It will be necessary for NSF to commence with an external review with a target completion in 2023 in order to evaluate the financial and programmatic viability of both proposed U.S. ELT projects, with the level of federal investment in at least one of the projects determined at the end of the review. Federal investment in either project should be predicated on:

1. Demonstration of financial viability with agreed-upon commitments from partners for all of the necessary capital and operations money, pending only NSF investment.
2. Final site selection in the case of the TMT.
3. A public share of telescope time (run through NSF's NOIRLab) roughly equivalent to the total federal investment of construction and operations expenses.
4. Full public archiving of all data taken by the ELTs, after a reasonable proprietary period. This applies to both federal and consortium telescope time.
5. Development of a management plan and governance structure for the joint project, agreed by all parties including the relevant observatory corporations and NSF.

Approval of the project is also subject to the recommendation in Section 5.1.1 that makes the initiation of any new astronomy MREFC project contingent on NSF developing a plan for managing the operations costs of the new facilities within its projected budget envelope.

Recommendation: The National Science Foundation (NSF) should conduct an external review of the U.S. extremely large telescopes, with a target completion date of 2023. If only one of the Giant Magellan Telescope or the Thirty Meter Telescope can meet the conditions enumerated above by the time of NSF's review, NSF should proceed with investment in that project alone.

Depending on the outcome, the decision rules for NSF are the following: In the case that only one project can proceed, NSF's investment of up to a 50 percent share in the project should be undertaken if doing so will ensure that the project has the financial resources to come to fruition. If NSF investment can only fund partnership in one telescope, but both are viable, NSF's investment should factor in complementarity to the ESO ELT, the ability to address the science questions of the Astro2020 survey, and the relative advantages of a larger diameter (D), which increases the sensitivity ~ D^2 to D^4 (depending on the science application), versus a larger field of view, which increases survey speed and the number of targets per observation.

7.6.1.3 CMB-S4

Observations of the CMB have not only been central to establishing the standard model of cosmology, but the telescopes designed to undertake them are becoming increasingly important for understanding phenomena ranging from transients to galactic ecosystems to the formation of cosmic structure. The advances possible with a new generation of receivers include searching for polarization signals from gravitational waves from the Big Bang and, when combined with Euclid, Roman, and Rubin Observatory, revealing a detailed picture of our cosmic web, its composition, and its evolution. At the same time, by tracing the electron pressure in halos of galaxies and galaxy clusters, CMB observations can trace feedback between the intergalactic medium, the circumgalactic medium, and the cores of galaxies.

Building on the scientific and technical progress brought about by decades of individual private and public investments by the U.S. community, we are poised in the next decade to make a major step forward in ground-based CMB studies. Over the last two decades, second- and third-generation ground-

based CMB experiments, deployed in Antarctica and Chile, have made significant advances, including detecting lensing B-mode signatures in the CMB, and the CMB-galaxy lensing cross power spectrum. The search for the tell-tale signature of cosmic inflation through its imprint on the B-mode polarization pattern of the CMB has pushed to fainter and fainter levels, disentangling foregrounds, and placing tighter constraints on this primordial signal. These observations have informed us how to analyze vast amounts of data and disentangle complex cosmological signals, and how to build theoretical models to extract parameters. The experiments have propelled progress by university groups and government labs to develop ever more sensitive, highly multiplexed bolometer detectors operating over a wide frequency range, and these efforts have informed the community how to design the next-generation facility to push these ground-based observations to their projected limit.

Realizing the ultimate scientific potential of ground-based CMB observations will take an effort far beyond what can be achieved simply by independently scaling up existing experiments. It will require a significant increase in the number of CMB detectors in operation, a wide range of independent frequency bands to separate out foreground contaminations, and it will require probing a combination of both large and small angular scales. While such an effort can be carried out using existing millimeter-wave observing sites in Chile and Antarctica, facilities at both must be carefully designed as part of a systemically planned program. Finally, while the United States has been the unrivaled leader in ground-based CMB observations, the needed project is of a scale that would benefit greatly from international participation in both scientific and technical aspects.

The Panel on Radio, Millimeter and Submillimeter Observations from the Ground (RMS) evaluated a number of CMB projects, and suggested that the CMB-S4 observatory as the compelling and timely next leap for ground-based observations. CMB-S4 is a joint effort of NSF and DOE that includes international participation. It will conduct a 7-year ultra-deep survey of a few percent of the sky from the South Pole with a combination of large and multiple small aperture telescopes observing from 30-270 GHz. This will be done in parallel with a 7-year deep/wide survey of roughly half the sky with additional telescopes sited in the Atacama desert in Chile. A TRACE analysis estimated the cost for design, development and construction to be $660 million (FY2020), within 15 percent of the project team's analysis and within uncertainties for this stage of development. CMB-S4 is well along in planning, and could achieve first-light as early as 2026-27. Although significant scale-up of the detector production is required, plans are in hand to accomplish this. Aerospace evaluated the project risk as medium-low.

This project engages the international cosmology communities, building upon the foundation of decades of ground- and space-based measurements of the CMB to take a major leap that will push CMB science to the next level. The scientific reach of this observatory goes well beyond cosmology. CMB-S4 will produce unprecedented maps of ~50 percent of the sky between wavelengths of 1 mm and 1 cm with a cadence that samples the entire area every other day, opening up discovery space and providing scientific data that will engage a broad swath of the astronomical community. Particularly compelling to the survey is the fact that these observations open the opportunity for systematic time-domain studies in this part of the electromagnetic spectrum for the first time.

Recommendation: The National Science Foundation and the Department of Energy should jointly pursue the design and implementation of the next generation ground-based cosmic microwave background experiment (CMB-S4).

Important to our recommendation is that CMB-S4 is a project with a balanced commitment from both NSF and DOE from inception, to design, implementation, operations and science. NSF nurtures and supports university groups with broad scientific and technical experience who have been leading ground-based CMB efforts both in Chile and in Antarctica, and that have been and will continue to train new generations of talent. DOE brings to bear the technical expertise of its national laboratories, scientific expertise including large scale computation, and importantly systematic management approaches that have proven to be effective for large-scale projects. The agencies have been working jointly and effectively to prepare for initiating this compelling project.

An important requirement for our strong endorsement is that the project broadly engage astronomers beyond the traditional CMB community. CMB-S4 will produce data sets of unprecedented sensitivity, cadence and spectral coverage that will advance general astrophysics and open discovery space opportunities for diverse scientific communities. Previous CMB experiments have not had the charge or funding to make data rapidly available and generally usable. It is essential that CMB-S4 produce transient alerts, as well as calibrated maps in all bands and on all angular scales that are openly usable and accessible on as rapid a cadence as practical. This is not necessarily at the same level of precision needed for CMB analysis. This will both maximize and justify the significant national investment in the observatory, even if it does require some nominal level of additional funding to accomplish.

7.6.1.4 The Next Generation Very Large Array

For the last four decades, the Karl Jansky Very Large Array (JVLA) Radio Telescope has been the premier observatory world-wide for accessing the sky at centimeter wavelengths. Likewise, the Very Long Baseline Array (VLBA), with its continental baseline, has extended centimeter radio observations to make images with exquisite angular resolution and perform precise astrometry. Both of these facilities have been upgraded since their inception, but they are now at a stage where further significant performance improvements are fundamentally limited by the quality of the antennae, and by their total number and allowable configurations.

The next generation Very Large Array (ngVLA) project is a powerful observatory that will replace both the JVLA and VLBA. The ngVLA is an array of up to 244 reflector antennas distributed across North America, operating at frequencies from 1.2 to 116 GHz. As conceived, it would achieve velocity resolution as fine as a fraction of a m/s, sub-milliarcsecond angular resolution, and high-fidelity imaging capabilities on scales from milliarcseconds to arcminutes. The project would have broad, flexible capabilities and provide science-ready data products accessible to a diverse community of users.

Such a facility would advance multiple high priority science questions from each of the six Science Panels,[14] and open discovery space. These include searching for diagnostic radio emission in compact object mergers from current and future ground- and space-based gravitational wave observatories, mapping the circumgalactic and intergalactic media, cold gas flows inside distant galaxies, and features on the surfaces of nearby stars. The ngVLA would resolve protoplanetary disks on scales more than 20 times finer than ALMA, potentially capturing images of planet formation in action. The ngVLA facility would be absolutely unique worldwide in both sensitivity and frequency coverage.

> **Conclusion:** It is of essential importance to astronomy that the JVLA and VLBA be replaced by an observatory that can achieve roughly an order of magnitude improvement in sensitivity compared to these facilities, with the ability to image radio sources on scales of arcminutes to fractions of a milliarcsecond.

A TRACE analysis of the ngVLA was performed, and the RMS panel undertook its own evaluation of the technical, cost and programmatic factors and risks. While there are some schedule threats related to antenna prototyping and the high required delivery rate, the technical risk is low, and the overall risk rating is medium-low. The TRACE budget assessment for design and construction is $3.2 billion (FY2020), which is within 5 percent of the RMS panel assessment, and ~20 percent higher than the project estimate. This discrepancy is reasonable given the early stage of development. The project aims to eventually secure 25 percent of the required funding from international partners, with 75 percent being provided by NSF. If design and prototyping were undertaken soon after the release of this report,

[14] See Table M.1 in the report of the Panel on Radio, Millimeter, and Submillimeter Observations from the Ground.

construction could start in 2027, with operations beginning with a partial array starting in 2034. In the survey's assessment, the biggest uncertainty is whether funding at the required level can be secured, given that this would be by far the largest project to be supported by the MREFC line.

Recommendation: The National Science Foundation (NSF) should proceed with a program to support science design, development, cost studies, and antenna prototyping for the Next Generation Very Large Array. After completion of the studies, NSF should convene a review to assess the project's readiness and available budget and proceed with construction if possible.

The project as presented to the survey is extremely ambitious in scale, and this could significantly extend the timeframe for commencement of science observations at design sensitivity. An important element of the design studies is whether the overall project cost can be significantly reduced with an acceptable impact on the science. For example, the survey notes that many of the important science objectives could be met with longer integration times, providing opportunity for rescoping of the hardware. Further, it will be important for the project to work during this time to secure international partners who would contribute to the project design, construction and operation at a significant level. This will not only reduce cost to NSF, increasing likelihood of timely completion, but tap into world-wide scientific and technical expertise.

The mid-decadal review will be a timely opportunity to examine the status of the intended design and design trades, and to assess them against the science goals of the survey. If sufficient progress is made on design, prototyping, and refinement of cost, and if budgets allow, an additional external review of the project would be necessary to consider whether to commence with implementation toward the end of the decade. If technical progress is not sufficiently rapid, or, if the required funding cannot be secured, the next decadal will need to weigh the implementation of the ngVLA relative to other opportunities.

7.6.2 Sustaining Activities: The Astronomy Mid-Scale Programs

Mid-scale programs across the entire range of scales (~$4 million–$120 million) are vital to the enabling foundation of astronomy research, and for capitalizing and amplifying return on our investment in major facilities. They enable new transformative capabilities by incentivizing creative approaches from the community for cutting-edge instruments and experiments. They ensure robust capabilities for basic research through continually refreshed instrumentation suites. They also provide broad access for the community across public-private partnerships, international system of platforms, observing modes, and wavelengths for individual-investigator initiated programs, large survey programs, and archival research.

As described in Chapter 6, in the last decade, NSF established a Mid-Scale Innovations Program (MSIP) within AST, and more recently, in 2018, an agency-wide Mid-Scale Research Infrastructure (MSRI) program. The survey received a large number of APC white papers for midscale projects, concentrated at the higher end of the cost range (~$100 million) that were evaluated by the OIR, Panel on Particle Astrophysics and Gravitation, and RMS program panels. All three panels provided multiple superb examples of compelling mid-scale ideas in this cost range. The panels all emphasize the high science value, cost effectiveness, and agility of mid-scale programs at all cost levels to address new science opportunities throughout the decade. Across the range of project scales, mid-scale programs are essential both for achieving the broad range of science prioritized by the survey, for addressing targeted strategic goals, and for ensuring that existing facilities have modern instrumentation to maximize their scientific productivity and community access (see Section 5.1.3).

Accordingly, the survey believes that the return from the MSIP and MSRI funding programs will be maximized if resources are deployed in a balanced manner that simultaneously accommodates:

- **Open competition of new ideas -** activating the community's creativity with minimal restrictions on scientific focus in order to fuel new, inventive, cutting-edge approaches that respond to emerging scientific opportunities.
- **Targeted solicitations designed to advance decadal priorities –** responding to identified scientific objectives that can be achieved using mid-scale facilities.
- **Opportunities targeted at sustaining and advancing instrumentation on existing telescopes –** maintaining U.S. competitiveness in ground-based astronomy, and optimizing scientific returns from current facilities.

Recommendation: The National Science Foundation (NSF) Division of Astronomical Sciences (AST) should create three tracks within the AST Mid-Scale Innovations Program and within (its share of) the NSF-wide Mid-Scale Research Infrastructure Program. The first track should be for regularly competed, open calls, the second track should solicit proposals in strategically identified priority areas, and the third should invite ideas for upgrading and developing new instrumentation on existing facilities. All tracks should solicit proposals broadly enough to ensure healthy competition.

This survey provides the following advice for each of the three tracks:

1. **Open calls** would continue to emphasize innovative ideas in any area of astrophysics over a wide range of project scales and scientific objectives, consistent with the approach taken in the current AST MSIP program.
2. The strategic priorities track is an essential addition to the existing mid-scale program structure to ensure that it is responsive to decadal and community strategic priorities. The survey expects that these strategic programs will be at the larger end of the mid-scale cost range (i.e., at the ~$100 million level). Therefore, partnerships with other organizations or agencies, including internationally, may be desirable or appropriate. Program directors would be empowered to weigh programmatic considerations in balance with the recommendations of external reviews. The survey has identified one top priority for this element, a time-domain astrophysics program, and two co-equal areas – highly multiplexed spectroscopy and radio instrumentation:
    a. **A time-domain astrophysics program.** This program would support a wide range of activities time domain astrophysics. A priority is to maximize the return from Rubin Observatory and other time domain facilities by, for example, supporting efforts to produce efficient triggers, perform time-domain data analysis, and optimize the identification, classification, and notification of transient events (often referred to as event brokers). It is essential that instrumentation aimed at effective time domain and multi-messenger follow-up and spectroscopy be supported. This element is the highest priority for immediate implementation of the mid-scale strategic areas. Based on white paper inputs, the expected costs for these efforts range from $4 million–$40 million.
    b. **Radio instrumentation**. The survey received compelling white papers that lay out exciting new projects in radio astronomy, including a wide-field radio camera, and projects to map the evolution of neutral atomic hydrogen in the very early universe. These are major, MSRI-2 scale efforts that could be competed for implementation starting this decade.
    c. **Highly multiplexed spectroscopy** – large surveys, such as that to be carried out by the Rubin Observatory, require extensive spectroscopic follow-up. Many of the science panels, as well as the OIR program panel, emphasized the need for new capabilities and especially those that are publicly available, to advance the survey's science priorities. Noteworthy science areas included galactic archeology and the spectroscopy of stars on a massive scale for understanding stellar abundances and evolution. In the near term, investments that provide public access to some combination of SDSS-V, DESI, and the

Subaru Prime Focus Spectrograph (PFS), or similar surveys, would help to advance science this decade with relatively modest funding, and later in the decade a major (MSRI-2 scale) investment could be made in a larger, dedicated facility.

3. **The sustaining instrumentation element** is intended to address the pressing need to maintain and upgrade capabilities on U.S.-led telescopes and to develop state of the art instrumentation on existing facilities to keep them at the scientific forefront. With the survey's top large recommendation being investment in the U.S. Extremely Large Telescope (ELT) program (Sec 7.6.1.1), the need for complementary instruments on a range of smaller OIR telescopes will become more pressing in the coming decade. Upgrades to 6–10 m class instrumentation will ensure the ability to conduct supporting and preparatory science. Smaller telescopes will be essential for conducting surveys, and will also serve as testbeds for demonstrating new technologies (Box 6.1). Sustaining instrumentation calls would be open to all facilities, public and private, and would support investments for private telescopes that emphasize community access in exchange for instrument investments (see Sec. 5.1.3 for an extensive discussion of this issue). In addition these calls would support upgrades to public facilities such as the Green Bank Telescope, Gemini, and CTIO.

External peer review remains the gold standard for recommendations and rankings in all three tracks. The selection criteria for all mid-scale projects would emphasize broad community access. This access could be gained through negotiated "dollars for community time" agreements (as in the former TSIP program), inter-facility "instrument time swap" agreements, public access to proprietary/consortium survey data, or in other ways. The guiding principle is that midscale investments serve to enhance the capacity of the portfolio of research capabilities to which the community has access.

Given the strong endorsement of many projects by the program panels, the analysis performed by the EF panel, the expected endorsement of ground-based solar physics projects by the solar and space physics decadal survey, and the survey's recommendation to add strategic calls to NSF's mid-scale programs, current mid-scale funding levels are inadequate. There is strong motivation to support projects across all scales, from $4 million to the large (>~$100 million) efforts, and across all wavebands. The survey estimates that this will require funding at a level of ~$50 million a year dedicated to AST, provided through a combination of MSIP and MSRI.

Conclusion: Current budget levels for AST mid-scale projects are not sufficient to advance the full range of astronomy and astrophysics priorities.

Recommendation: The National Science Foundation should increase the funding available in its mid-scale programs that support astronomy and astrophysics with a target of reaching $50 million per year for the combination of the Mid-Scale Innovations Program and the Mid-Scale Research Infrastructure Program.

The appropriate distribution of funding among the three tracks is best determined by proposal pressure. All elements of the program are essential to the survey's objectives.

7.6.3 NSF Physics Projects Central to Astro2020

As discussed in Chapter 1, the task of the survey is to "develop a comprehensive research strategy to advance the frontiers of astronomy and astrophysics." Increasingly facilities and projects that are planned and supported through the Division of Physics at NSF are essential to advancing these frontiers. The LIGO gravitational wave observatory, part of NSF's Gravitational Physics Program, is the prime example. The discovery of gravitational waves from merging black holes in 2015 propelled LIGO to its current essential position as a premier observatory for understanding the demographics and astrophysical

implications of black holes, and for identifying the sources of heavy elements in the universe. These are among the most rapidly advancing areas in modern astrophysics, and future discoveries are likely to bring additional surprises. Therefore, the future of gravitational wave detection is central to progress in astronomy and astrophysics, and to this report. Another facility with strong overlap with astrophysics is IceCube, which is part of NSF Physics Astroparticle Physics program. New messengers and new physics is a central theme of Astro2020, and firm associations of high-energy neutrino events with astrophysical objects promises to provide unique information on particle acceleration in some of the most extreme environments near black holes.

NSF recognizes that central motivations for future investments in gravitational wave detection and high energy neutrino detection lie in astronomy and astrophysics. In their briefing to the steering committee in July, 2019, NSF emphasized that its Physics Division would like to have an evaluation of the importance of these, and other programs in NSF Physics (in the divisions of Plasma Physics, Nuclear Physics, and Elementary Particle Physics), but that ranking them relative to projects led out of AST would not be helpful, given the different advisory and funding mechanisms. Further, the survey was only given budget guidance for NSF AST, and for the agency-wide MREFC program. Of the NSF Physics Division programs, gravitational wave and high energy neutrino detection stand out for having essential scientific motivation in astrophysics.

7.6.3.1 Technology Development for Future Ground-based Gravitational Wave Observatories

Gravitational wave astrophysics is one of the most exciting frontiers in science. One of the survey's key science priorities is the opening of new windows on the dynamic universe, with gravitational wave detection at the forefront. To achieve this goal, the continued growth in sensitivity of current-generation facilities such as LIGO through phased upgrades is essential. In 2018, the "A+" upgrade was approved to reduce the quantum and thermal noise in the detectors. Installation of this upgrade is beginning, and it is expected to begin operating in 2024, with an ultimate astrophysical reach of about a billion light years for neutron star coalescences, although it may take several years to achieve that sensitivity. New technologies are being developed for more advanced detectors in the current LIGO facilities (named "Voyager") that would bring additional sensitivity improvements, including the use of silicon rather than fused silica for the test masses and suspension fibers, operation at 100 Kelvin, and a laser source of a different wavelength.

In the longer term, the international community is planning for next-generation interferometers, such as the U.S. Cosmic Explorer, and the European Einstein Telescope that will make dramatic leaps in science capability. The rate of binary neutron star detections will be sufficient to make precise measurements of the Hubble constant through the detection of electromagnetic counterparts. For merging black holes the signals will be loud enough for precision tests of general relativity, and for nearby neutron star coalescences tight constraints can be placed on the equation of state of dense material. With these facilities black hole mergers can be detected out to high redshifts. In addition, intermediate mass black hole mergers can be used to probe their cosmic evolution and provide insights into the first seeds.

Both concepts for third generation gravitational wave detectors, Einstein Telescope and Cosmic Explorer (CE), have longer baselines to diminish the impact of seismic and thermal noise, requiring large costs in vacuum systems. The CE concept is a single L shaped detector constructed on Earth's surface with arms 40km long, using technology being developed for the A+ upgrade of the 4 km, for a 10-fold increase in sensitivity. Similar to strategic mission technologies, maturation research is needed for more economical vacuum systems, use of heavier masses, improved vertical seismic isolation and reduction of gravitational influence of nearby terrain ("Newtonian noise"). The use of Voyager technology (cryogenics, different materials and wavelengths) could be used in the future to further improve CE sensitivity.

Conclusion: Gravitational wave detection is an essential capability for advancing the frontiers of astronomy and astrophysics. Next-generation ground-based gravitational wave observatories can achieve breakthrough capabilities and accomplish science central to the priority objectives of this survey. Continuous technology development will be needed this decade for next generation detectors like Cosmic Explorer. These developments will also be of benefit to the astrophysical reach of current facilities.

While not funded out of AST, these efforts are central to achieving the science vision laid out in the survey's roadmap, and the survey strongly endorses their importance to astronomy and astrophysics.

7.6.3.2 The IceCube Generation-2 Neutrino Observatory

Observations of high-energy neutrinos enable astrophysical advances in the study of some of the most energetic phenomena in the universe. In particular, the most extreme accelerators in the universe produce huge luminosities of charged particles and accompanying gamma rays and neutrinos, with per-particle energies ranging up to the TeV-PeV range, and sometimes higher. The IceCube observations of the diffuse neutrino flux suggest a dominant population of sources that are gamma-ray obscured, showing that neutrino observations are essential for understanding and studying such energetic phenomena.

A large-scale MREFC investment by NSF in IceCube-Gen2 would greatly enhance this observatory's capabilities. "IceCube-Gen2 will increase the annual rate of observed cosmic neutrinos by a factor of ten compared to IceCube, and will be able to detect sources five times fainter than its predecessor. Furthermore, through the addition of a radio array, IceCube-Gen2 will extend the energy range by several orders of magnitude compared to IceCube."[15] The primary scientific objectives for this upgrade are to resolve the bright, hard-spectrum TeV-PeV diffuse neutrino background into discrete sources, make the first detections at higher neutrino energies, and identify neutrino emission with specific astrophysical sources in order to gain insight into sites of extreme particle acceleration. The PAG panel, supported by a TRACE study of the observatory upgrade, finds that the project is well-understood, uses mature technology, and with a cost of $345 million in FY2020 is feasible to implement this decade. This survey was not charged to make project recommendations to NSF PHY; however, we endorse the observatory as important to key astrophysics scientific objectives of this survey.

Conclusion: The IceCube-Generation 2 neutrino observatory would provide significantly enhanced capabilities for detecting high-energy neutrinos, including the ability to resolve the bright, hard-spectrum TeV-PeV neutrino background into discrete sources. Its capabilities are important for achieving key scientific objectives of this survey.

7.6.4 Prioritization of NSF Sustaining and Frontier Activities

For NSF, the survey's top priority in the medium and large category is to complete the observatories in development, and ensure they are fully supported for operations and science (Table 7.1). The mid-scale programs, MSIP and MSRI are recommended not just to be maintained, but for a significant augmentation. For new projects, the survey does not rank the sustaining mid-scale program augmentation against major new facilities. Both are essential for an optimal, balanced program. Among the large, frontier AST projects, the survey gives the U.S. ELT program the highest priority due to the large scientific reach, and the pressing funding needs and maturity of the two constituent telescope

[15] The IceCube-Gen2 Collaboration: M.G. Aartsen, R. Abbasi, M. Ackerman, J. Adams, J.A. Aguilar, M. Ahlers, M. Ahrens, et al., 2019, IceCube-Gen2: The window to the extreme universe, white paper submitted to the Astro2020 Decadal Survey, arXiv:2008.04323.

projects. The survey does not prioritize between CMB-S4 and the ngVLA design and development; both will provide transformative science in different arenas, and the efforts should both proceed as soon as funding becomes available. In the sustaining programs category the augmentation of mid-scale programs and their restructuring to include strategic calls is the single priority. The funding distribution between open, strategic, and sustaining instrumentation calls needs to be balanced, and adjusted over the decade to respond to proposal pressure and strategic needs. At the request of the agency, the survey does not rank projects led out of NSF Physics. However, the survey strongly endorses the central role played by ground-based gravitational wave observatories to many of the survey's high-priority science questions, and urges NSF to invest in a healthy program to develop technologies for future LIGO upgrades and next-generation facilities. In the frontier observatory category the survey concludes that NSF Physics Division's IceCube-Gen2 neutrino observatory will have impact on several of the priority science questions and has a central role in the New Messengers, New Physics theme, but again it is not directly ranked.

7.7 NASA'S PROGRAM OF RECORD

In this section we provide an assessment of, and advice where needed on the implementation plans for Roman, Athena, and LISA, because these missions are still at a stage in development where this advice could impact the Astro2020 scientific agenda.

7.7.1 Roman Space Telescope

The Roman Space Telescope (formerly WFIRST) was the highest space-based priority of the Astro2010 decadal survey. It was envisioned in that report to be a $1.6 billion, 1.5 m near-IR telescope enabling diffraction-limited imaging and low-resolution spectroscopy over a wide field of view. The primary science drivers included: (1) constraining dark energy through measurements of weak gravitational lensing, supernovae distances, and baryon acoustic oscillations, and (2) statistically assessing the frequency of Earth-mass planets on orbits of ~1 AU and greater by carrying out a microlensing survey towards the bulge of the Milky Way Galaxy. Astro2010 additionally called out an open "guest observer" program that would take advantage of the large field of view. A key aspect of the mission's recommendation was its relatively low technical and cost risks. The Astro2010 plan called for a 5-year mission primarily focused on survey science, with an additional 5-years to "improve statistical results" and to "further broaden the science program."

The current design for the Roman Space Telescope has many similarities to that originally envisioned by Astro2010, but also has key differences. The National Reconnaissance Office gave NASA a 2.4 m space telescope that became the centerpiece of WFIRST-AFTA (for Astrophysics Focused Telescope Assets). Coincident with the change to a larger telescope, an exoplanet-imaging Coronograph Instrument (CGI) was also added. The primary goal was for technology demonstration, although the initial design had significant scientific capability. Since then, to maintain overall project schedule and budget constraints (NASA has adopted a $3.5 billion cost ceiling for Roman), CGI has had its capabilities significantly descoped. In addition to its wide field of view, a key feature of Roman is its rapid time to slew and settle, with no Earth occultations given its orbit at L2.[16] The telescope is currently scheduled for a 2026 launch.

[16] L2, or Lagrange point 2, is a location in space directly behind Earth as viewed from the Sun where gravitational forces are balanced such that a telescope placed at this point will stay in line with Earth as Earth moves around the Sun.

A number of reviews were undertaken to evaluate the consistency of the new mission design with Astro2010 recommendations. The 2014 NRC WFIRST/AFTA[17] study concluded that the increase in aperture resulted in a powerful mission meeting all of the science goals of Astro2010, and that the addition of CGI was a positive step if it focused on technology development and did not drive key mission requirements. In order to remain consistent with the balanced program recommended by Astro2010, the WFIRST/AFTA report also noted containing costs would be critical, and it recommended an independent review of mission scope prior to formal adoption. The resulting WFIRST Independent External Technical/Management/Cost Review (WIETR) was held in 2017, and in 2020 the mission underwent a successful Key Decision Point-C review, formally confirming the implementation phase.

Evaluating Roman in the current landscape, the mission remains both powerful and necessary for achieving the scientific goals set by Astro2010. Roman's cosmological constraints complement those of Euclid and Rubin, with its main contribution expected to come from the $1.8<z<2.5$ redshift range. At lower redshifts its constraints on the expansion history are not expected to improve upon Euclid's, owing to that telescope's much wider sky coverage. As the systematic errors of Euclid, Rubin, and Roman are different, the three experiments will provide important verification of each other's results; this is particularly important if Euclid finds significant deviations from standard models for Dark Energy. Roman is also the only platform in the coming decades that can produce a statistical census of planetary occurrence as a function of orbital separation and mass, from terrestrials to gas giants, beyond 1 AU. Such a survey would "pick up" where Kepler completeness falls off, just inside of ~1 AU, although (like Kepler) the actual sensitivity to true Earth analogs in the habitable zone is relatively low.

> **Finding:** The Roman Space Telescope remains both powerful and necessary for achieving the scientific goals set by *New Worlds, New Horizons* (Astro2010). It will carry out cosmological measurements complementing those of Euclid and Rubin Observatory, and Roman's microlensing survey will probe planetary occurrence over orbital separations not constrained by Kepler or TESS.

Roman also provides substantial scientific capabilities that will contribute to achieving the science vision presented in Astro2020. The Astro2020 Science Panel reports describe in detail where Roman's capabilities will provide significant advances relative to the key questions and discovery areas. Out of thirty questions and discovery areas, Roman will directly impact fourteen. Although the most obvious advances will be in cosmology and exoplanets, Roman's immense discovery potential beyond those areas almost ensures that its highest impact results will come from other, and possibly unforeseen, directions. Examples are studies of high redshift galaxies, active galactic nuclei, dark matter in Local Group dwarf galaxies, stellar populations in galaxies in the local volume, the stellar mass function in star clusters, and optical/near-IR counterparts to gravitational wave events. More generally Roman will be a premier facility for obtaining deep, high resolution imaging and slitless spectroscopy at optical/near-IR wavelengths, with a field of view that is a factor of 200 larger than HST.

The scientific landscape has changed significantly since Roman was first recommended by Astro2010. However, with the change to a larger telescope, Roman has also become more capable. Compared to the situation in 2010, Roman is now just one of multiple "Stage IV" projects (as defined by the Dark Energy Task Force), including the Vera Rubin Observatory's Legacy Survey of Space and Time, Euclid, and the Dark Energy Spectroscopic Instrument. As a result of the discovery of gravitational wave sources in 2015, and the burgeoning of time-domain astronomy this decade, Astro2020 identified "New Windows on the Dynamic Universe" as one of its priority science areas for the coming decades. Roman, with its wide field of view, and flexible pointing could provide unique time domain surveys, possibly coordinated with other efforts.

---

[17] National Research Council, 2014, *Evaluation of the Implementation of WFIRST/AFTA in the Context of New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, Washington, D.C.

In light of the altered landscape and new opportunities, it is reasonable to ask whether the allocations of survey time recommended by the Science Definition Team in mid-2015 are still optimal. The planned imaging, spectroscopy, and microlensing surveys continue to be essential parts of Roman's mission, and a wealth of science will result from their data products, far beyond the measurement of Dark Energy parameters and exoplanet statistics. However, as currently planned, the balance of these surveys with the equally promising GO program may not be ideal, when evaluated in light of the updated survey scientific objectives. It is beyond the scope of this survey to recommend an appropriate rebalancing of time, however given that there are still 3 or 4 years until launch, a dedicated re-evaluation of the scientific program in light of this survey's scientific priorities is warranted.

Conclusion: The scientific landscape and the Roman Space Telescope's capabilities have changed significantly since it was first envisioned by *New Worlds*, *New Horizons*, and the currently planned balance of surveys and guest investigator-led observations may not be optimally suited to take advantage of new scientific opportunities.

Recommendation: NASA Astrophysics Division should hold a non-advocate review of the Roman Space Telescope's science program to set the appropriate mix of survey time devoted to the weak lensing, baryon acoustic oscillations, supernovae, and microlensing programs relative to guest investigator-led observing programs during the primary 5 year mission.

7.7.2 The Athena X-ray Observatory

NASA has joined as a partner in the second of ESA's Flagship Cosmic Visions missions (L2), Athena. This high energy observatory, currently scheduled for launch in 2031, is oriented towards science themes of the "Hot and Energetic Universe." Its science instruments enable wide-field X-ray imaging and sensitive spatially-resolved spectroscopy of X-ray-emitting objects. NASA plans to invest ~$200 million to $300 million, split roughly equally between hardware contributions for half the amount, and establishment of a U.S. Guest Observer (GO) program and U.S. Data Center. The planned hardware contributions include components of the two science instruments, use of the X-ray and Cryogenic Facility (XRCF) at Marshall Space Flight Center (MSFC), and a Soft-Ride system to dampen launch vibrations. These contributions leverage unique U.S. capabilities and facilities.

Athena's science instruments map well onto a wide range of the priority science questions identified by Astro2020—19 out of the 30 science questions/discovery areas will be directly addressed by the mission. The newly-established NASA Project Office, and appointment of U.S. members to Athena's science working groups, keep the U.S. community engaged in the project. Plans for a U.S. Guest Observer program and U.S. Data Center will ensure that U.S.-based scientists will be well-supported in analysis of Athena observations. When Athena begins operations in the early 2030's it will be the premier X-ray observatory in space, and the United States will be well-positioned to play a significant role in the science it produces.

Conclusion: The scope of the U.S. investment in the European Space Agency's Athena mission is appropriate both for the hardware contribution as well as for the U.S. Guest Observer and science center. This investment will enable substantial scientific involvement by the U.S. community in this exciting mission.

7.7.3 The Laser Interferometer Space Antenna Mission

The Laser Interferometer Space Antenna (LISA) was the third highest ranked large strategic mission in Astro2010, after WFIRST (now Roman) and a major augmentation to the Explorer program. At the time that LISA was evaluated by Astro2010, enthusiasm for the science that this low-frequency gravitational wave mission could achieve was very high, however Astro2010 judged that advancement to the highest priority large strategic mission should be contingent on the success of the LISA Pathfinder technology demonstration mission, and also further development of the mission concept, costs and risks. The mission's scientific potential was judged to be at the very highest level, but more technological development and risk reduction were deemed necessary prior to recommending LISA for a mission start in the 2010–2020 decade.

Since Astro2010 there has been major progress on both scientific and technical fronts. The LISA Pathfinder mission demonstrated crucial, high-risk components of the mission's precision metrology capability, exceeding all of its performance requirements. Pathfinder placed two test masses in a near-perfect gravitational free fall, and it controlled and measured their motion with unprecedented accuracy. To do this it used inertial sensors, a laser metrology system, a drag-free control system and an ultra-precise micro-propulsion system. This demonstrated LISA's highest risk components in a space environment and with a practical implementation. LIGO's detection of gravitational wave sources in 2015 was a transformational event that reinvigorated excitement about LISA's scientific potential.

In 2017 ESA accepted a proposal to develop a version of LISA with launch expected 2034 or after. Based on LISA's promise and the enthusiasm of the U.S. scientific community, NASA established the NASA LISA Studies Office (NLSO) to estimate the cost of, and coordinate U.S. contributions to the mission. NASA currently plans to contribute an equivalent of $400 million in mission hardware by supplying the telescope, laser, and charge management systems as well as phasemeters and micro-thrusters. As LISA develops, additional areas where NASA can make a critical or especially effective contribution may become apparent.

The survey committee evaluated the scientific case for LISA in light of recent LIGO/Virgo measurements, and with the priority science questions in mind. The scientific case for the mission remains rich and compelling. For example, LISA will observe hundreds of stellar mass binary black hole systems, some of which would cross into the band of ground-based observatories like LIGO/Virgo weeks to months later. LISA also complements nano-Hz gravitational wave measurements using pulsar timing arrays; the latter are sensitive to billion solar mass black hole mergers while LISA is sensitive to $\sim 10^{4\text{-}6}$ solar mass black hole mergers (Figure 2.14). Several target sources for LISA will produce electromagnetic signals that can be followed up in several different bands with space and ground observatories. Some events may produce particles detectable on Earth. LISA will thus greatly advance our multi-messenger view of the universe.

While LISA builds on experience gained from ground-based gravitational wave detectors, it will have distinct data analysis and processing challenges. Unlike LIGO/Virgo, at any instant LISA will measure the superposition of multiple sources, with many signals lasting months or years. The instrument noise (statistical and systematic) will also be qualitatively different than that of ground-based detectors. The LISA analysis will therefore differ significantly from that used for LIGO-like instruments, and new computational techniques must be developed in order to interpret the data.

The format of the LISA analysis program within ESA has not been set. Various models are under consideration. There is an opportunity for significant U.S. engagement and coordination with ESA to continue and extend the collaboration begun with the hardware through to the data analysis. With anticipated contributions from NASA, LISA's sensitivity (the L3 proposal to ESA)[18] will be close to that of the Astro2010 reference mission[19] and Astro2010 mid-decadal assessment. In regards to scientific

[18] Amaro-Seoane et al., 2017, arXiv 1702.00786.
[19] Stebbins et al, LISA Astro2010 RFI #2 Space Response

partnership and data analysis, there are many future opportunities for the U.S. community for NASA to embrace.

> **Conclusion:** ESA's LISA mission remains a very high priority for the U.S. community, and NASA contributions of hardware and data analysis tools are essential to ensuring the full scientific capability of the mission is achieved, as envisioned by *New Worlds*, *New Horizons* and the subsequent mid-decadal assessment. It is also essential to maintain a vibrant U.S. community to prepare for data analysis and science.
>
> **Recommendation: NASA should work with the European Space Agency to ensure the Laser Interferometer Space Antenna (LISA) achieves the full scientific capability envisioned by *New Worlds, New Horizons*. NASA should continue calls for LISA Preparatory Science with a known cadence during the decade. After a jointly developed plan for LISA data analysis and management are clear, and a few years prior to launch, NASA should establish funding for LISA science at a level that ensures U.S. scientists can fully participate in LISA analysis, interpretation, and theory.**

7.8 BUDGETARY ANALYSIS

This section evaluates the budgetary requirements for the recommended program separately for NASA, NSF, and the DOE. The evaluation adopts the cost and schedule profiles determined by the TRACE analyses, with minor adjustments made in some cases to reflect the judgement of the program panel regarding technology and programmatic readiness. It is important to note that this analysis assumes that the agencies are able to provide the optimal funding profiles for the given project or mission. It further assumes that the agencies make the recommended early investments in project/mission and technology maturation. Deviations from the optimal budget profiles that reduce funding in years of peak spending will extend development periods and increase the total mission cost relative to this analysis. This survey emphasizes the need for investment in maturation programs (e.g. the Great Observatories Mission and Technology Maturation Program for NASA, and ngVLA design and prototyping efforts for NSF) so that costs and schedules can be more accurately determined prior to mission/project adoption. This will assist NASA and NSF in their planning to meet peak costs successfully, an important factor in containing total project costs.

7.8.1 NSF Analysis

7.8.1.1 MREFC Program

The budget profile analysis shown in Figure 7.8 presents the program outlined in the roadmap of new ground-based major projects in terms of the NSF share of expected program/project cost, and it compares the total cost with the budget projection provided by NSF. The chart runs through FY2041 to capture the expected completion of the ngVLA. The TRACE cost and schedule estimates are used for construction, and project or Program Panel estimates for operations. In some cases, the phasing and durations were adjusted to manage several factors: approximate budget, technology development/readiness, and other programmatic factors. For example, the U.S. ELT program, consisting of TMT and GMT, was spread out over an additional 2 years, lowering the peak spending proposed by the projects, but consistent with the OIR program panel's judgement related to the rapidity with which NSF funding could be provided. For the ngVLA, the RMS panel recommended an additional 2 years of design and development relative to plans provided by the project prior to any major ramp up of construction efforts, and we incorporate that into this analysis. The NSF MREFC share assumed for the ELTs is 25

percent of the total construction costs for each telescope, for CMB-S4 40 percent of the costs are assumed to be borne by NSF (60 percent by the DOE), and for the ngVLA, the NSF share is assumed to be 75 percent of total costs, with 25 percent to be identified in the future international partners. These fractional funding levels were adopted from the project white papers and presentations. The MREFC budget profile also includes current commitments and a growing wedge for the agency-wide MSRI mid-scale programs as provided to us by NSF. Mid-scale projects are discussed elsewhere, but the committee assumed approximately 15 percent of the NSF-wide MSRI budget line shown here might be successful AST projects. The remainder of the mid-scale funding would need to come out of the AST budget to achieve the total $50 million a year target.

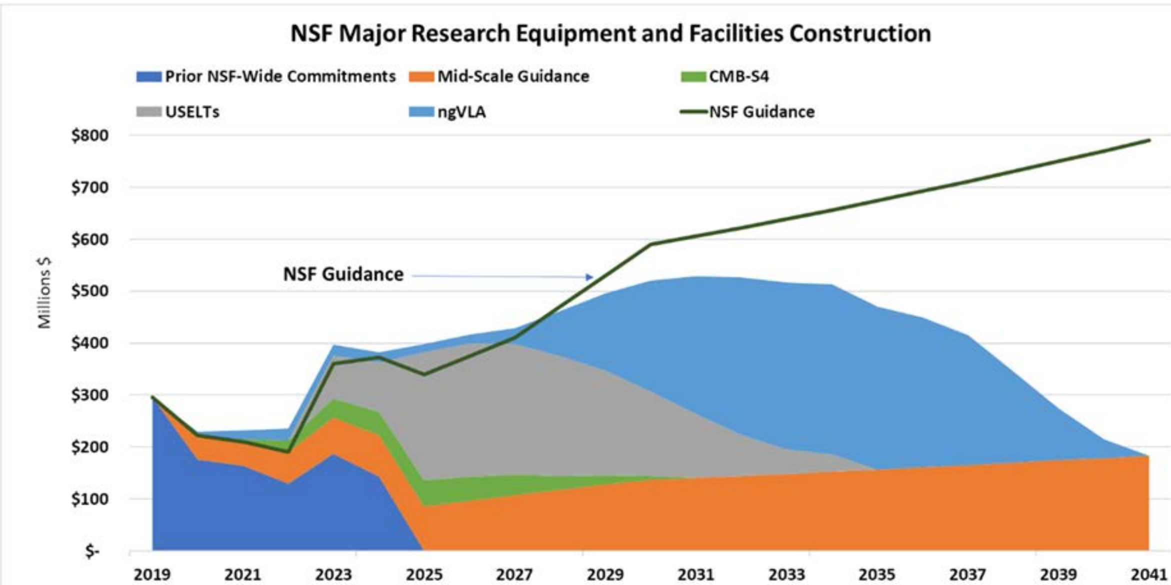


FIGURE 7.8 Recommended program for the National Science Foundation (NSF) MREFC. The chart assumes that the agency-wide midscale MSRI funding wedge given to the survey by NSF is realized. It also includes agency-wide prior commitments, and the new NSF construction funding required to realize the large AST projects recommended by the survey. The solid line shows the MREFC budget guidance provided to the survey by NSF through 2030, and extrapolated beyond this using 2.7% inflation. Note that for CMB-S4, U.S. ELTs, and the ngVLA, which have additional contributions from other agencies and partners, only the NSF-share of the total funding is shown. The operations costs for new facilities are included in the budget chart for NSF AST; see Figure 7.9.

7.8.1.2 NSF AST Budget

Budget projections for NSF AST shown in Figure 7.9 take into account the existing components of the budget in the areas of education, research, and infrastructure, as well as additions recommended from this decadal survey. The specific recommended items in chapters 3, 4, 6, and 7 for the division span all three of these categories; only recommendations for which specific enough guidance is given to be able to assign a dollar amount are included in these projections. The infrastructure component, in particular, encompasses operations and maintenance for facilities including the National Solar Observatory, the National Radio Astronomy Observatory, the National OIR Astronomy Research Laboratory, and others, as well as an AST Portfolio Review Implementation, Midscale Research Infrastructure, and Research Resources. The starting point for projecting the NSF AST budget and

examining impacts is the FY2019 budget actuals.[20] Operations costs associated with new facilities expected to come online in this decade and the next also figure into the calculations. DKIST has already begun initial operations, and the budget ramps up to full operations in 2022. The Rubin Observatory is gearing up for science operations estimated to $30 million in 2024. Operations budgets associated with new MREFC projects are phased in at the appropriate time as new demands on the budget; operations for the JVLA ramp down as the ngVLA ramps up, in accordance with project and panel guidance. Operations for ngVLA also phase in a partial array for limited early science while the remaining antennas are being integrated, based on advice from the RMS panel.

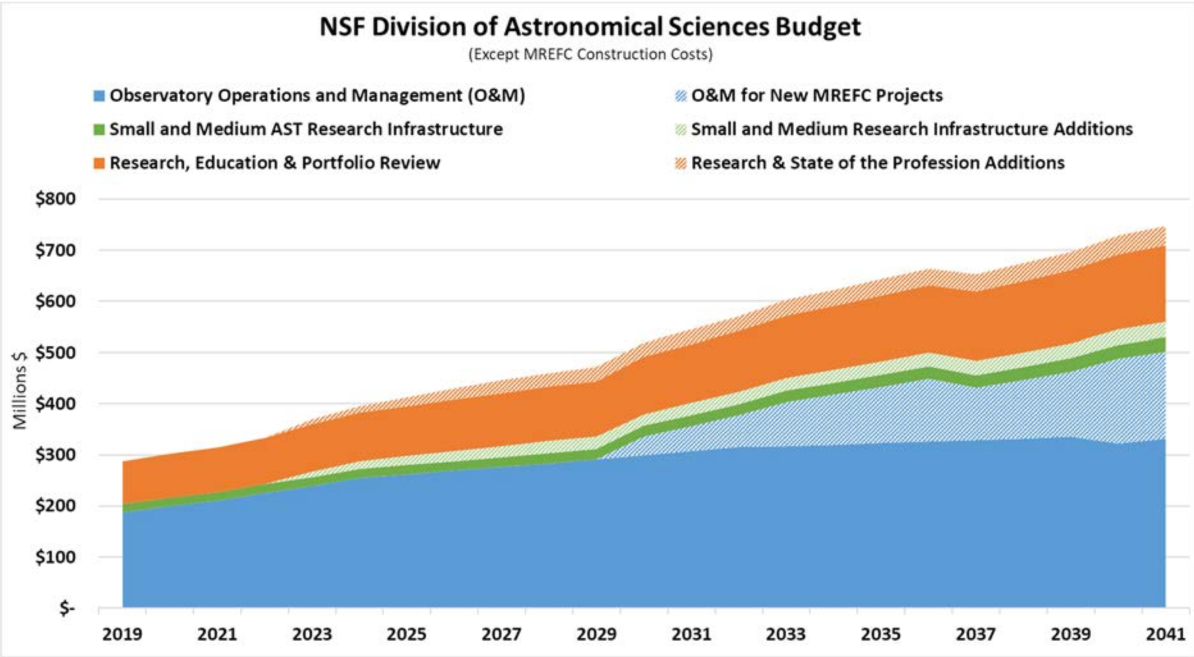


FIGURE 7.9 NSF AST budget requirements for implementing the programs recommended by this decadal survey. The required funding is based on existing budget items as well as the recommended additions. Solid colors indicate existing budget components, adjusted for inflation at 2.7 percent. The orange bar indicates research, along with education, research resources, and other minor budget components. Hatched shading indicates the additions to these three areas resulting from recommendations in this report. Small and Medium Research Infrastructure Additions is in large part the ATI augmentation.

Figure 7.9 shows the required budget profile for NSF AST that results from these additions. The increases in non-operations funding have only a modest impact on the total budget profile for the division. Adding operations costs for new MREFC-funded facilities on top of the existing budget burden for operations and maintenance clearly becomes unsustainable going forward, without recourse to a new paradigm for accounting for operations impact. There will need to be significant increases in research funding to accommodate the demand based on impending science from new facilities. The increase to individual investigator grants called out in chapter 4 and included here is a minimum amount. Even with a new paradigm for operations funding that is not within NSF AST, the growth in the field spurred by new facilities will need to be matched by similar increases in support for research to ensure a robust science environment.

[20] https://www.nsf.gov/about/budget/fy2021/pdf/27_fy2021.pdf

7.8.2 NASA Analysis

The budget profile shown in Figure 7.10 presents the roadmap outlined above in terms of the expected program or project cost, and then compares the total with the optimistic budget projection provided by NASA. The chart runs through FY2050 to capture the expected completion of the IR/O/UV large strategic mission. The phasing of each program element in the roadmap is adjusted in order to manage several factors: available budget, technology development/readiness, pre-cursor programmatic work and studies and science priorities established by the survey committee. Though the IR/O/UV mission drives the total program to exceed the yearly available budget between FY2035 and FY2043, the integrated cost of all elements in the program through FY2043 is approximately $23.5 billion, about $0.15 billion less than the integrated available budget over the same period. As demonstrated in the past, it is expected that NASA will work within the federal budgeting process to assure that peak budgetary requirements are met while sustaining its portfolio as a whole.

In Figure 7.10, the Great Observatory Mission Maturation and Technology Development (GOMMTD) program is shown broken into its constituent parts. This program consolidates mission maturation activities, technology development and management activities. All large strategic mission activities start within the maturation program. When a large strategic mission achieves sufficient maturity, and has a scope consistent with decadal recommendations, mission-specific funding begins. In parallel, mission maturation and technology development for additional large strategic mission commence. In Figure 7.10, a notional future (deep blue) $5 billion class GO is shown undergoing the same development strategy as implemented for the IR/O/UV mission.

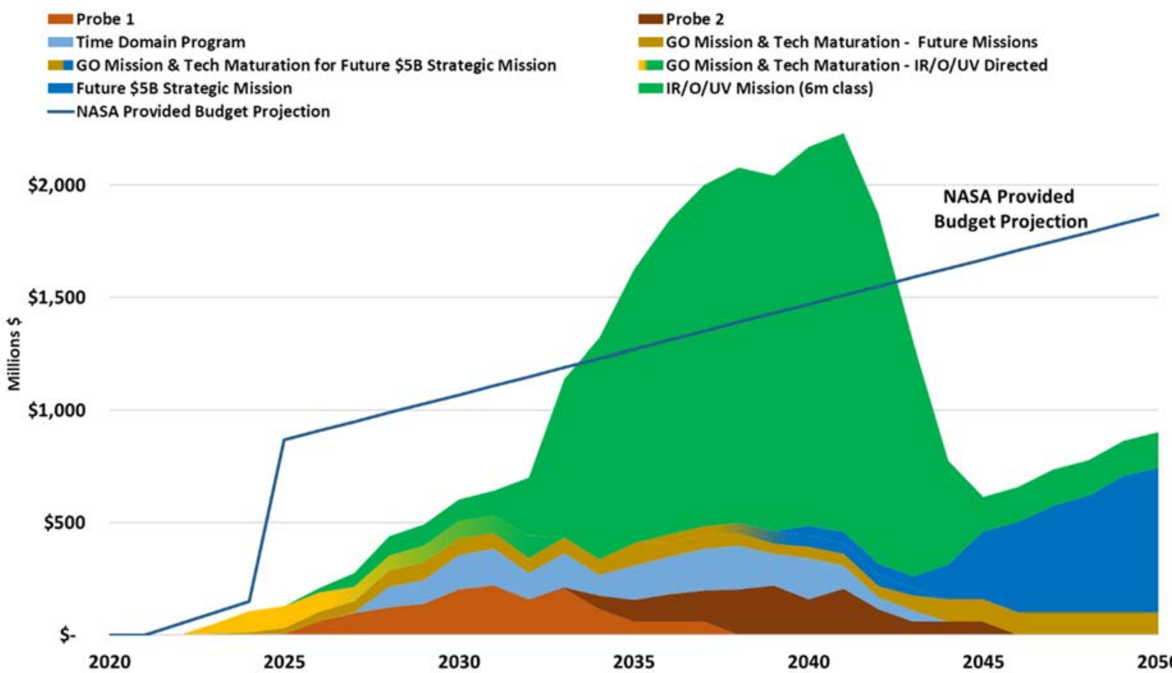


FIGURE 7.10 Astro2020 recommended program for NASA. This budget profile shows approximate funding requirements associated with construction and operation of all space-based medium and large recommendations. For the IR/O/UV mission, operations are assumed to extend beyond 2050. The ultimate project/program profiles and budget requirements will depend on the actual implementation and on NASA's budgeting process. The chart shows a program whose costs integrated through FY2043 are approximately equal to the budget available over the same period. The solid line indicates the optimistic budget projection that NASA provided to the survey. The jump in NASA's available astrophysics budget around 2025 reflects completion of Roman and reduction in other current commitments.

7.8.3 DOE Analysis

The CMB-S4 observatory is the only new facility where the project assumed a formal commitment of funding from the DOE. Adopting the total cost estimate from the TRACE analysis, the 60 percent DOE share for construction results in a total commitment of $408 million. The operations costs are also assumed at 60 percent for DOE and 40 percent for NSF.

7.9 ANALYSIS OF CONSISTENCY WITH BUDGETARY GUIDANCE

The survey committee performed an analysis to assess whether the proposed program of new activities presented in the roadmap is consistent with envisioned budget profiles provided by the agencies. All three agencies urged the survey committee to present an ambitious vision that would motivate increased federal investment in their astronomy and astrophysics portfolios, and we use the optimistic scenarios given to us for planning. The TRACE project cost and schedule estimates for construction have been adopted where available, with some elements shifted in time to reflect technical readiness and other factors. Details of these assumptions are provided in the Program Panel appendixes for the individual projects and activities. Integrated over time, the proposed portfolios fit within the agency budget guidance, even though in years of peak spending for major projects, this guidance is exceeded. The survey considers this to be appropriate for agencies with budget lines that are driven by the requirements of major projects.

Appendixes

A

Statement of Task and Panel Descriptions

STATEMENT OF TASK

The National Academies of Sciences, Engineering, and Medicine shall convene an ad hoc survey committee and supporting study panels to carry out a decadal survey in astronomy and astrophysics. The study will generate consensus recommendations to implement a comprehensive strategy and vision for a decade of transformative science at the frontiers of astronomy and astrophysics. The committee, with inputs from study panels covering the breadth of astronomy and astrophysics, will carry out the following tasks:

1. Provide an overview of the current state of astronomy and astrophysics science, and technology research in support of that science, with connections to other scientific areas where appropriate;
2. Identify the most compelling science challenges and frontiers in astronomy and astrophysics, which shall motivate the committee's strategy for the future;
3. Develop a comprehensive research strategy to advance the frontiers of astronomy and astrophysics for the period 2022-2032 that will include identifying, recommending, and ranking the highest priority research activities—taking into account for each activity the scientific case, international and private landscape, timing, cost category and cost risk, as well as technical readiness, technical risk, and opportunities for partnerships. The strategy should be balanced, by considering large, medium, and small activities for both ground and space. (Activities include any project, telescope, facility, experiment, mission, or research program of sufficient scope to be identified separately in the final report.) For each recommended activity the committee will lay out the principal science objectives and activity capabilities, including assumed or recommended activity lifetime, where possible;
4. Utilize and recommend decision rules, where appropriate, for the comprehensive research strategy that can accommodate significant but reasonable deviations in the projected budget or changes in urgency precipitated by new discoveries or unanticipated competitive activities;
5. Assess the state of the profession, using information available externally and, if necessary, data gathered by the study itself, including workforce and demographic issues in the field. Identify areas of concern and importance to the community raised by this assessment in service of the future vitality and capability of the astronomy and astrophysics work force. Where possible, provide specific, actionable and practical recommendations to the agencies and community to address these areas. This report shall be made available following the completion of the study.

PANEL DESCRIPTIONS—SCIENCE PANELS

Panel on Compact Objects and Energetic Phenomena

The Panel on Compact Objects and Energetic Phenomena will identify and articulate the scientific themes that will define the frontier in research of compact objects and energetic phenomena in the 2022-2032 decade. Its scope will include white dwarfs; neutron stars; pulsars; magnetars; stellar mass black holes; compact binary systems; novae; supernovae: gamma-ray bursts; fast radio bursts; physical processes and accretion onto supermassive black holes; and gravitational radiation and high-energy particles and radiation from astrophysical sources.

Panel on Cosmology

The Panel on Cosmology will identify and articulate the scientific themes that will define the frontier in cosmology research in the 2022-2032 decade. Its scope will include the early universe, the cosmic microwave background, cosmological tests and parameters, the epoch of reionization observations and theory large scale structure, dark energy, dark matter (excluding direct detection), and gravitational lensing and microlensing as applied to cosmology, as well as astrophysical tests of fundamental physics.

Panel on Galaxies

The Panel on Galaxies will identify and articulate the scientific themes that will define the frontier in galaxy research in the 2022-2032 decade. Its scope will include observations, theory, and simulations of galaxy formation and evolution, galactic structure and dynamics, galaxy clusters, stellar populations in galaxies, the intergalactic medium, chemical evolution of galaxies, gravitational lensing as applied to galaxy structure, demographics of supermassive black holes, co-evolution of galaxies and supermassive black holes, relevant aspects of Milky Way science, and related phenomena associated with active galactic nuclei.

Panel on Exoplanets, Astrobiology, and the Solar System

The Panel on Exoplanets, Astrobiology, and the Solar System will identify and articulate the scientific themes that will define the frontier in research of exoplanets, astrobiology, and the solar system in the 2022-2032 decade. Its scope will include the detection, demographics, and physical characteristics of exoplanets, solar system observations relevant to Astro2020, astrobiology, stellar phenomena and activity that impact detectability and characterization of exoplanets, and effects of stellar activity on the evolution and habitability of planets. The panel will consider as inputs the congressionally mandated reports *Exoplanet Science Strategy* and *An Astrobiology Strategy for the Search for Life in the Universe*.

Panel on the Interstellar Medium and Star and Planet Formation

The Panel on the Interstellar Medium and Star and Planet Formation will identify and articulate the scientific themes that will define the frontier in research of the interstellar medium and the formation of stars and planets in the 2022-2032 decade. Its scope will include the interstellar medium and star formation in the Milky Way and nearby galaxies, astrochemistry, interstellar plasmas, protoplanetary disks and debris disks, and planet formation.

Panel on Stars, the Sun, and Stellar Populations

The Panel on Stars, the Sun, and Stellar Populations will identify and articulate the scientific themes that will define the frontier in research of stars, stellar populations, and the Sun in the 2022-2032 decade. Its scope will include stellar structure and evolution, stellar activity and variability, brown dwarfs, solar astronomy as relevant to Astro2020, resolved stellar populations and star clusters in the local group, and stellar nucleosynthesis and chemical evolution.

PANEL DESCRIPTIONS—PROGRAM PANELS

Panel on an Enabling Foundation for Research

The Panel on An Enabling Foundation for Research will summarize the current state of resources and support, identify major challenges, and make suggestions to the Astro2020 committee on the topics of theory, computation, and simulation; data collection, archiving, and analysis; facilities, funding, and programs; laboratory astrophysics; and general technology development. The Panel's suggestions will be incorporated into a program for all of astronomy and astrophysics by the Committee on Astro2020.

Panel on Electromagnetic Observations from Space 1

The Panel on Electromagnetic Observations from Space 1 (EOS1) will identify and suggest to the decadal survey committee a program of federal investment in research activities that involve observations of astrophysical phenomena primarily by means of ultraviolet, optical, and near-infrared electromagnetic measurements from space. The EOS1 panel will also consider technology development needs to support the program. In formulating its conclusions, the EOS1 panel will draw on several sources of information: (1) the science forefronts identified by the Astro2020 science panels, (2) input from the proponents of research activities, and (3) independent risk, technical readiness, and cost evaluations. The EOS1 panel's suggestions will be integrated into a program for all of astronomy and astrophysics by the Astro2020 Committee.

Panel on Electromagnetic Observations from Space 2

The Panel on Electromagnetic Observations from Space 2 (EOS2) will identify and suggest to the decadal survey committee a program of federal investment in research activities that involve observations of astrophysical phenomena primarily by means of radio, far-infrared, and high-energy electromagnetic observations from space; and research activities that involve gravitational radiation or particle detection from space. The EOS2 panel will also consider technology development needs to support the program. In formulating its conclusions, the EOS2 panel will draw on several sources of information: (1) the science forefronts identified by the Astro2020 science panels, (2) input from the proponents of research activities, and (3) independent risk, technical readiness, and cost evaluations. The EOS2 panel's suggestions will be integrated into a program for all of astronomy and astrophysics by the Astro2020 Committee.

Panel on Optical and Infrared Observations from the Ground

The Panel on Optical and Infrared Observations from the Ground (OIR) will identify and suggest to the decadal survey committee a program of federal investment in ground-based research activities that involve observations of astrophysical phenomena primarily by means of optical and infrared measurements from the ground. The OIR panel will also consider technology development needs to

support the program. In formulating its conclusions, the OIR panel will draw on several sources of information: (1) the science forefronts identified by the Astro2020 science panels, (2) input from the proponents of research activities, and (3) independent risk, technical readiness, and cost evaluations. The OIR panel's suggestions will be integrated into a program for all of astronomy and astrophysics by the Astro2020 Committee.

Panel on Particle Astrophysics and Gravitation

The Panel on Particle Astrophysics and Gravitation (PAG) will identify and suggest to the decadal survey committee a program of federal investment in research activities exploring areas at the interface of physics and astronomy such as gravitational radiation, gamma-ray astronomy, cosmic rays, and neutrinos. The PAG panel will also consider technology development needs to support the program. In formulating its conclusions, the PAG panel will draw on several sources of information: (1) the science forefronts identified by the Astro2020 science panels, (2) input from the proponents of research activities, and (3) independent risk, technical readiness, and cost evaluations. The PAG panel's suggestions will be integrated into a program for all of astronomy and astrophysics by the Astro2020 Committee.

Panel on Radio, Millimeter, and Submillimeter Observations from the Ground

The Panel on Radio, Millimeter and Submillimeter (RMS) Observations from the Ground will identify and suggest to the decadal survey committee a program of federal investment in ground-based research activities that primarily operate in the radio, millimeter, and submillimeter portions of the electromagnetic spectrum. The RMS panel will also consider technology development needs to support the program. In formulating its conclusions, the RMS panel will draw on several sources of information: (1) the science forefronts identified by the Astro2020 science panels, (2) input from the proponents of research activities, and (3) independent risk, technical readiness, and cost evaluations. The RMS panel's suggestions will be integrated into a program for all of astronomy and astrophysics by the Committee on Astro2020.

PANEL DESCRIPTION—PANEL ON STATE OF THE PROFESSION AND SOCIETAL IMPACTS

The Panel on State of the Profession and Societal Impacts will gather information on the health and demographics of the astronomy and astrophysics community and make actionable suggestions to the Astro2020 committee on the topics of demographics, diversity and inclusion, workplace climate, workforce development, education, public outreach, and relevant areas of astronomy and public policy. The panel's suggestions will be incorporated into a program for all of astronomy and astrophysics by the Committee on Astro2020.

ADDITIONAL INFORMATION FROM THE STEERING COMMITTEE TO THE SCIENCE PANELS

Astro2020's steering committee gave additional instructions to the science panels as they carried out their work of defining science themes for the next decade. Regarding content, the science panels were asked to provide a brief review of the current state of the science in their topic areas, and determine four important science questions to be addressed in the next decade and one area that shows great promise for discovery. The science panels were also asked to provide a summary of the general capabilities needed to address the science questions and discovery areas without going into an excess of quantitative detail.

Linking general capabilities to specific ground or space activities in consideration by the decadal survey was out of scope for the science panels. To accomplish their work, the panels relied on the 572 science white papers submitted by the astronomy and astrophysics community, the expertise of the panel members, discussion with other panels about common topics, and invited speakers if necessary to fill in any gaps. Structurally, the science panels were requested to keep their reports brief, keeping figures and citations to a minimum, but they were not constrained in how they chose to organize their writing.

ADDITIONAL INFORMATION FROM THE STEERING COMMITTEE TO THE PROGRAM PANELS

Astro2020's steering committee gave additional instructions to the program panels as they carried out their work of suggesting programs of federal investment in research activities for the next decade. These programs were then to be integrated by the steering committee into a recommended program for all of astronomy and astrophysics. Regarding content, the program panels were asked to provide a brief summary of the current state of the relevant program area to include facilities and programs in development, international capabilities, and major gaps in capabilities available to the U.S. astrophysics community. The program panels were also charged to assess the ability of current and proposed projects under consideration to address the science panels' questions and discovery areas, to comment on the technical, risk, and cost evaluations (TRACE) of the proposed projects, and to identify key areas of technical development or precursor research activities. The program panels were requested to identify the projects that were best suited for and most readily prepared to address the science priorities in the next decade; however, they were not asked to provide a ranked list of investments. The program panels had the option, but were not required, to discuss how the proposed projects might be modified to better fit the science priorities or the expected future budget availability. For medium-scale projects usually chosen by competitive selection, the program panels were asked to give examples of these projects and to discuss the balance of medium- and small-scale activities versus larger strategic investments. Detailed discussion of additional research and theory programs was primarily the task of the Panel on an Enabling Foundation for Research.

To accomplish their work, the panels relied on 294 activity, project, and state of the profession (APC) white papers submitted by the astronomy and astrophysics community, the expertise of the panel members, independent TRACE analyses of selected projects, and as needed, invited speakers and additional requests for information from the proposed projects. Structurally, the program panels were requested to keep their reports brief but were not constrained in how they chose to organize their writing.

B

Report of the Panel on Compact Objects and Energetic Phenomena

INTRODUCTION

The stars shining in the sky have been familiar since the dawn of humanity, but their fates as they end their lives is a story understood only gradually over the past century. Compact objects—white dwarfs, neutron stars, and black holes—are the remnant cores of ordinary stars after their nuclear burning ends. These exotic objects are characterized by extremes of gravity and are often a source of high-energy radiation and particles. White dwarfs (WDs), the cores of the lightest stars, have masses similar to the Sun, but in a volume a million times smaller (the size of Earth). Neutron stars (NSs), the collapsed cores of some massive stars, again have masses similar to the Sun, but are only the size of a city, resulting in an extreme density comparable to atomic nuclei. Rapidly rotating and highly magnetized NSs, called pulsars—some spinning hundreds of times per second—emit regular pulses of radiation like exceptionally stable cosmic lighthouses. Magnetars are NSs that are so extremely magnetized that their magnetic fields can tear the NS crust and cause violent starquakes.

Other massive stars leave behind stellar-mass black holes (BHs), gravitational singularities in the spacetime of general relativity. Once viewed as a speculative hypothesis, their existence has been firmly established in multiple ways, with masses ranging from a few to many tens of solar masses ($M_\odot$). In addition to stellar-mass BHs, supermassive black holes (SMBHs) in the 10^5–10^{10} $M_\odot$ range are observed at the centers of many galaxies. Accretion of matter onto these SMBHs is understood to power emission from active galactic nuclei (AGN) and quasars, the most luminous objects in the universe. The existence of BHs between these extremes (intermediate-mass black holes, or IMBHs) remains an open question.

Many compact objects are members of binary star systems. If the binary companion is an ordinary star, then stellar and binary evolution can allow mass transfer and accretion from the donor companion onto the compact object, often mediated by an accretion disk. Accretion onto compact objects is an efficient power source for radiation, leading to systems called cataclysmic variables (accreting WDs) and X-ray binaries (accreting NSs and BHs). Until recently, X-ray binaries provided the only means to observe stellar-mass BHs. However, stellar evolution in an accreting binary eventually causes the donor companion to form a second compact object (a process that the binary itself may or may not survive). Binaries consisting of a pair of compact objects can produce strong gravitational wave (GW) emission, resulting in angular momentum loss that eventually leads to coalescence (merger) of the binary and a GW transient. The successful detection of GWs from compact binary mergers in the past few years has opened a profoundly powerful new window on the study of compact object systems.

Some compact objects are sources of relativistic jets—collimated outflows of matter accelerated to nearly the speed of light. Compact objects are also closely linked to supernovae (SNe), gamma-ray bursts (GRBs), classical novae, and other explosive transients. Energetic phenomena associated with compact objects manifest through multiple messengers, including electromagnetic (EM) radiation ranging from low-frequency radio waves to the highest-energy gamma rays, GW radiation, high-energy neutrinos, and perhaps ultra-high-energy cosmic rays (UHECRs).

The past decade has seen extraordinary progress in the study of compact objects and energetic phenomena. Several of the major observational breakthroughs had been theoretically predicted in advance, including:

- The direct detection of GWs from merging compact binaries.
- The discovery of kilonovae (a new type of transient arising from merging NSs) and their associated ultra-relativistic jet outflows, through simultaneous detection of GWs and EM counterparts.
- The first imaging of a BH shadow, through radio interferometry of the SMBH in the nearby galaxy M87.

Other key breakthroughs included:

- The first detection of astrophysical high-energy neutrinos.
- The first discoveries of NSs more massive than 2 $M_{\odot}$
- The first discoveries of stellar-mass BHs substantially more massive than 15 $M_{\odot}$ (the pre-2010 record).
- The discovery of ultraluminous X-ray pulsars (NSs apparently accreting matter at a rate many hundreds of times larger than the spherical Eddington limit).
- The detection of large numbers of tidal disruption events, and the emergence of other important classes of astrophysical transients including fast radio bursts and superluminous supernovae.

These observational breakthroughs have been matched by major advances in theoretical calculations and modeling. Advanced numerical simulations have also enabled substantial progress in understanding the physical processes that govern accretion disks, relativistic jets, particle acceleration, supernovae, stellar evolution, and compact binary coalescence.

Primed by this recent progress and attentive to the expectation of powerful new capabilities in the coming decades, this report is organized around four key science questions and one outstanding discovery area. Future progress in all these areas share two foundational needs. The first is support for a broad range of theoretical and computational studies, as well as next-generation computing facilities for multidimensional radiation hydrodynamic and particle-in-cell simulations and numerical relativity. The second is support for the next generation of observatories, public data access, public data products, and tools for data mining.

B-Q1. WHAT ARE THE MASS AND SPIN DISTRIBUTIONS OF NEUTRON STARS AND STELLAR BLACK HOLES?

Among the measurable properties of compact objects, two fundamental quantities are their mass and spin, which can constrain their birth and evolution. Significant advances in measuring these quantities have been made recently, and more are expected in the coming decade. In particular, the GW detections of NS and BH binary mergers have opened new avenues for measuring masses and spins. Precision NS mass measurements are now possible for binary radio pulsars beyond NS-NS systems, and recent work has extended the NS mass range above 2 $M_{\odot}$, approaching the theoretical upper mass limit near ~2.5 $M_{\odot}$. GW detections of merging BH binaries have similarly proven the existence of >20 $M_{\odot}$ BHs merging to form >40 $M_{\odot}$ BHs, all substantially heavier than the BH population observed in X-ray binaries of the Milky Way Galaxy. Simultaneous measurements of NS mass and radius (or, equivalently, tidal deformability) through both X-ray pulsar timing and GW detection of NS-NS mergers has been achieved, pointing the way to eventual measurement of the full NS mass-radius relation and the equation of state of

cold matter at supranuclear densities. New X-ray continuum surveys will more completely map the X-ray binary populations in the Milky Way and beyond, while next-generation radio surveys should find thousands of new radio pulsars, including binary systems where mass measurements are possible. Massive optical/IR surveys for stellar radial velocity and photometric variations, astrometric motion, and microlensing will begin to probe the huge, previously unexplored population of isolated, free-floating or noninteracting compact object/stellar binary systems. These efforts will directly inform the next generation of theoretical calculations of binary star evolution, simulations of massive star SN collapses and explosions, NS and BH formation, and GW-driven mergers.

B-Q1a. What Do the Mass and Spin Distributions Tell Us About Neutron Star and Black Hole Formation and Evolution?

Explaining the observed birth-mass and birth-spin distributions of NSs and BHs remains an important unsolved problem. A detailed understanding of the NS mass and spin distributions could map to different massive star progenitors; map to different binary evolution channels; or distinguish between iron-core collapse, accretion-induced collapse, and electron-capture SNe. For example, the lowest mass NSs are currently expected to form from the lowest-mass collapsing massive stars. Similarly, in some models of SNe, the lowest-mass BHs should be ~5 $M_\odot$. Whether or not there is a true “mass gap” in the distribution of NS and BH birth masses between 2.5 and 5 $M_\odot$ constrains both the SN explosion mechanism and the potential for “fall back” during or after the explosion, with direct connection to the chemical enrichment of iron-peak elements in galaxies over cosmic time. (Note the new announcement in June 2020 of the compact binary merger GW190814, one of whose progenitor components was a 2.6 $M_\odot$ compact object of unknown type: either a massive NS or a light BH.)

As with NSs, the BH mass and spin distributions trace the physical origin of isolated and binary BHs. The GW discovery of binary BHs with component masses >20 $M_\odot$ was a major surprise to many astrophysicists, even though it had been predicted. The BH mass distribution, combined with the distribution and orientation of spins, may help reveal the physical origin of these BHs: do they arise from normal massive stellar binary evolution, triple/multiple star systems, or dynamical scattering in very dense stellar systems like globular clusters? That there might be gaps or breaks in the mass distribution at even higher masses may be understood as arising from known and hypothesized evolutionary pathways of single massive stars. For example, the theory of pair-instability SNe predicts the existence of a gap in the BH mass distribution in roughly the 50–140 $M_\odot$ range. (Note the new announcement in September 2020 of the BH-BH merger GW190521: the masses of its product and one of its progenitors lie within, or near the edges of, the pair-instability gap.)

The rich array of questions that the panel is poised to explore includes the following: Are there features in the NS and BH mass distributions? Is there a sharp cutoff or a gradual decline in the BH mass function at low and high mass? What is the distribution of binary mass ratios? Is there a significant population of NS-BH binaries, or of second-generation, hierarchically formed massive BHs? The question of whether most BHs are born spinning rapidly further constrains possible formation mechanisms. The observation of a large sample of BH-BH mergers out to high redshift may reveal how these distributions evolve with redshift and metallicity.

B-Q1b. What Is the Population of Noninteracting or Isolated Neutron Stars and Stellar-Mass Black Holes?

More than 100 million NSs and perhaps more than 10 million BHs exist in the Milky Way Galaxy. Yet, our knowledge of these systems is mostly confined to pulsars and accreting systems, with just a handful of nonpulsar/nonaccreting systems known. This may lead to a substantial bias in our

understanding of the underlying population. Can we detect the huge expected population of free-floating BHs and "quiet" NSs? Can we identify a substantial population of massive dark companions in binaries? This represents a huge discovery space over the next decade and beyond. Because most massive stars occur in binary or higher-order multiple systems, the mass distribution of NSs and BHs constrains critically uncertain aspects of interacting binary and massive star evolution. A substantially larger sample of NS- or BH-stellar binary systems could potentially allow for an understanding of the relative rate of SN success/failure as a function of metallicity because the companion stars can be characterized in detail.

B-Q1c. What Is the Equation of State of Ultradense Matter?

A more complete characterization of the NS and BH mass and spin distributions will also reveal fundamental physics. The equation of state of matter at supranuclear density cannot be probed in terrestrial laboratories, but it manifests through the NS mass-radius relation and is thus often referred to as the NS equation of state (NS-EOS). Constraints on the NS-EOS come from populating regions of the mass-radius plane with measurements from NSs, and from seeking out the extremes of the NS and BH mass distributions. In particular, pulsar searches have provided the most massive NSs, limiting the maximum NS mass from below, whereas the discovery and characterization of low-mass BHs limit the maximum NS mass from above. Double pulsar systems may provide the first measurement of a NS moment of inertia, providing a complementary constraint. The highest-spin NSs can, in principle, also constrain the NS-EOS through centrifugal limits; however, for reasons still not understood, the pulsars found to date all have spin frequencies significantly below the theoretical maximum.

Another approach to constrain the NS-EOS is to directly and simultaneously measure the masses and radii of a large sample of NSs. Precise radius measurements are particularly challenging. X-ray pulse profile modeling of NSs with the Neutron Star Interior Composition Explorer (NICER) and the Laser Interferometer Gravitational-Wave Observatory (LIGO)/Virgo detection of the NS-NS merger GW170817 have both demonstrated this approach and provided preliminary constraints on the NS-EOS. A large population of GW-detected NS-NS mergers will yield important independent constraints on the maximum NS mass and the tidal deformability (and hence the associated radius), and thereby a strong constraint on the NS-EOS. More sensitive X-ray timing of a large sample of millisecond pulsars (both rotation-powered and accretion-powered) may allow for a complete characterization of the NS mass–radius plane, mapping out the full NS mass-radius relation rather than relying on a few, isolated mass-radius points.

Relevant Measurements and Capabilities

Over the next decade and beyond, our understanding of the NS and BH mass and spin distributions will be revolutionized by new observations across the EM spectrum and by a host of other messengers. In particular, currently planned high-frequency (Hz/kHz) GW detectors will directly reveal the mass and spin distributions of compact objects in merging BH-BH binaries to $z \sim 1$ and NS-NS binaries to ~1 Gpc. More advanced GW detectors could extend this range to $z \sim 20$ (BH-BH) and $z \sim 5$ (NS-NS), respectively, yielding an enormous sample of compact objects that would provide fundamental constraints on general relativity, the ultradense matter equation of state, the diverse formation channels of merging BH and NS binaries, and the evolution of these channels with redshift and metallicity. Achieving this requires an order of magnitude improvement in the sensitivity of ground-based detectors beyond current design sensitivity, with even greater sensitivity improvement at lower frequencies (down to a few Hz). Multiple detectors will enable the localization of thousands of NS-NS binaries to better than 1 deg^2, well matched to synoptic EM surveys for prompt counterparts.

The next generation of wide-field pulsar searches will reveal new extreme pulsar systems, pulsar–WD binaries, more double-pulsar binaries, and perhaps the first pulsar-BH binary, all providing

fundamental mass and spin measurements. However, this will require pulsar search/timing instrumentation for both single-dish (using multipixel receivers) and array observatories (where pulsar searches are extremely computationally expensive). It will also require observational investment for follow-up timing of each new pulsar to determine if they are scientifically interesting. Larger-area X-ray timing observatories can directly constrain the masses and radii of NSs in both accreting and bursting NS systems through pulse profile modeling, while more sensitive X-ray imaging and spectroscopy can more fully reveal the X-ray binary populations of our and other galaxies and allow for disk reflection line measurements of BH spins.

Missions like Gaia and its precision astrometric successors will probe the large population of stellar binaries with dark compact-object companions. Combined with current and forthcoming massive spectroscopic surveys like the Dark Energy Spectroscopic Instrument (DESI), Sloan Digital Sky Survey (SDSS)-V, and the next generation of massively multiplexed stellar spectroscopy, the very large (but still uncharacterized) population of NS- and BH-stellar binaries of the Milky Way will be revealed. The population of free-floating and otherwise undetectable NSs and BHs will be explored for the first time with the Wide-Field Infrared Survey Telescope (WFIRST) gravitational microlensing survey. Follow-up observations with 8 m-class telescopes and 30 m-class telescopes (extremely large telescopes, or ELTs) will be required for more complete characterization of individual events. The Laser Interferometer Space Antenna (LISA) and other mHz GW detectors will open up the currently inaccessible regime of the numerous quiescent and otherwise EM-dark GW lighthouses of NS and BH binaries that are destined to merge on Gyr time scales in our own galaxy.

B-Q2. WHAT POWERS THE DIVERSITY OF EXPLOSIVE PHENOMENA ACROSS THE ELECTROMAGNETIC SPECTRUM?

Astrophysical transients (energetic events that appear in the sky only briefly) are signposts of the most catastrophic events in spacetime. Known classes include stellar explosions, stellar disruptions by supermassive BHs, stellar eruptions, and mergers of stars or compact objects, as well as very short bursts of radio emission with uncertain origin. Transients lie at the intersection of several critical areas of modern astrophysics and cosmology. Stellar explosions create dust, help trigger the formation of new stars, and produce BHs and NSs. Some transients are valuable “standard candles” used to trace the acceleration of the universe. The death throes of massive stars deposit radiative and mechanical energy into the interstellar medium (ISM) and drive the chemical enrichment and evolution of their host galaxies. Shocks from massive stellar explosions provide a key way to constrain the still-mysterious mass-loss history of massive stars before core-collapse. Fast radio bursts offer a completely new probe of the intergalactic medium (IGM) and large-scale structure, while NS mergers play an important role in the synthesis of the heaviest elements of the periodic table. Additionally, shocks launched by a variety of astronomical transients constitute unique laboratories for relativistic particle acceleration under extreme physical conditions.

Recent technological advances have led to a revolution in the investigative power of astronomical time-domain surveys, which in turn have led to the discovery of new classes of transients (e.g., fast radio bursts, superluminous SNe, stellar mergers), and enabled the exploration of new parameter spaces. Upcoming optical surveys like the Legacy Survey of Space and Time (LSST) with the Vera C. Rubin Observatory will take this effort to the next stage, and this revolution will soon encompass a broader range of wavelengths outside the optical and gamma-ray bands—for example, radio with Square Kilometer Array precursors; X-rays with Extended Roentgen Survey with an Imaging Telescope Array (eROSITA); and near-infrared (near-IR) with WFIRST. Additionally, the recent discovery of GWs and light from the NS merger event GW170817, and the possible association of the high-energy neutrino event IceCube-170922A with blazar TXS 0506+056, clearly demonstrate the rich connection between EM transients and other astronomical messengers. The time is ripe to fully realize the discovery potential

of this multimessenger data stream and develop a complete physical understanding of the rich phenomenology of the transients that are observed.

B-Q2a. When and How Are Transients Powered by Neutron Stars or Black Holes?

Understanding the central engines (newly formed compact objects like magnetars and BHs) that power many explosive transients continues to be a fundamental astrophysical challenge. For example, superluminous SNe (SLSNe) have peak luminosities ~10–100 × those of normal core-collapse SNe, requiring an additional source of energy beyond the traditional neutrino-powered SN mechanism. Possibilities include rotational energy from magnetars, or gravitational binding energy liberated during BH accretion. Central engines can also manifest through relativistic collimated outflows (i.e., jets in long-duration-GRBs, tidal disruption events, or compact binary mergers). Which unique physical conditions enable some transients to launch ultra-relativistic jets? What are the nature and properties (e.g., mass, spin) of the newly formed compact objects? Is there EM emission before compact binary mergers, and what can it tell us about the properties of the progenitor systems? Current speculation is that fast radio bursts are also manifestations of cataclysmic events involving NSs or BHs. What actually triggers fast radio bursts? (The panel notes new observations in 2020 associating some fast radio bursts with magnetars.)

B-Q2b. When and How Are Transients Powered by Shocks?

Transient mass ejections produce shocks, either with an external medium or internal to the outflow itself. Shocks accelerate particles that subsequently radiate photons through thermal and nonthermal processes, and allow for an efficient conversion of shock kinetic energy into radiation. For which transient phenomena does the efficient conversion of kinetic shock energy to radiation represent the dominant source of energy, and why? What determines the radiative efficiency of thermal and nonthermal processes?

Remarkably, classical novae (thermonuclear outbursts from accreting WDs) have recently been discovered to produce detectable gamma rays, something not theoretically predicted. This finding suggests an unexpectedly important role for strong shocks in nova phenomenology. Observations and better understanding of these nearby common transients may help test the hypothesis that SLSNe and other stellar explosions are also shock-powered. In stellar explosions, the breakout of shock radiation is the very first EM signal that reaches the observer, and it carries a wealth of information about the exploding star. What can we learn about the largely unconstrained population of SN progenitors from systematic observations of shock breakouts across the EM spectrum?

B-Q2c. When and How Are Transients Powered by Radioactivity?

The radioactive decay of newly formed nucleosynthetic products is a known source of energy powering the optical light-curves of both ordinary core-collapse (type II) and thermonuclear (type Ia) SNe. The thermalization of gamma rays originating from the β-decay of ^{56}Ni in the ejecta of type Ia SNe provides a key energy input to their early light-curves. While it is clear the type-Ia SNe originate from carbon-oxygen WDs in binaries, the mechanism that triggers the explosion and the mapping between type-Ia SN observables and progenitor types remain unclear. Is the progenitor a WD-WD merger in a double-degenerate binary, or an accreting WD in a single-degenerate binary? Although type-Ia SNe were traditionally understood to arise from a WD exceeding the 1.4 $M_{\odot}$ Chandrasekhar limit, recent theoretical and observational progress has highlighted that WDs over a relatively large range of masses can explode as a type-Ia SN, including both sub- and super-Chandrasekhar progenitors (the latter arising either from

rapid rotation or from a WD-WD merger). Multiple channels may be possible. A critical unknown in type-Ia SN studies is to determine what fraction of explosions is produced by each channel.

Recent observations demonstrate that heavy elements were produced through *r*-process nucleosynthesis in the neutron-rich ejecta of the NS merger event GW170817, and that the subsequent radioactive decay powered a transient known as a kilonova, which evolved on a week-long time scale. These observations establish NS mergers as one of the sites of *r*-process nucleosynthesis. After this landmark discovery, the frontier is now to answer the question: Are NS mergers the *main site* of *r*-process nucleosynthesis and the *main source* of heavy chemical elements in the universe, or are there supernovae and other core-collapse events that contribute significantly to the heavy *r*-process budget of the universe?

B-Q2d. What Are the Unexplored Frontiers in Transient Phenomena?

During the past decade, we have witnessed the proliferation of discoveries of unanticipated classes of transients with observed properties that challenge traditional classification schemes or paradigms. The most prominent examples are fast radio bursts and gamma-ray transients associated with classical nova outbursts. Other examples include peculiar thermonuclear SNe (e.g., type Iax), stellar eruptions preceding core collapse, transients in the luminosity gap between classical novae and SNe, the very rapid time scales of fast and blue optical transients (FBOTs), and the extreme luminosities of SLSNe.

At the same time, there are solid theoretical predictions of astrophysical transients that are still lacking an uncontroversial observational example. These include the accretion-induced collapse of a WD into a NS, pair-instability SNe, and the merger of a NS-BH binary. Which other transient classes have yet to be revealed? Which theoretical models will find observational confirmation? What is the role of other sources of energy (like magnetic reconnection, free neutron decay, recombination) to power astrophysical transients? How do the explosions of the first stars appear?

Relevant Measurements and Capabilities

Progress in the field of astrophysical transients critically depends on two key capabilities: discovery power and understanding. Discovery power is effectively enabled by observing facilities with large fields of view, able to monitor the sky in real time. A healthy ecosystem of optical/infrared transient surveys with a range of sensitivities and temporal cadences, combined with prompt public release of discoveries and data, is required to find and characterize transients over the entire range of time scales, distances, and luminosities produced by the cosmos. Wide-field monitors in the ultraviolet, X-ray, and low/medium-energy gamma-ray bands are needed to open the fields of SN shock breakout, to enable systematic exploration of tidal disruption events, and to maintain detection capabilities of GRBs in conjunction with GW events. Wide-field MeV gamma-ray spectroscopy is needed to detect nuclear line emission from SNe. High-frequency (Hz/kHz) GW observations with better localization and sensitivity are needed to enable larger samples of NS mergers with EM counterparts. An MeV neutrino observatory with an order-of-magnitude improvement in sensitivity relative to hyper-Kamiokande would allow detection of ~1 core-collapse SN per year. A wide-field radio time-domain survey with arcsecond-localization capabilities (like the Canadian Hydrogen Intensity Mapping Experiment [CHIME], but with sufficiently precise positioning for multiwavelength follow-up) is needed to map the radio transient sky and enable fast radio burst detection and characterization.

To understand the physics powering these transients, a variety of follow-up observing machines is required. Massively multiplexed optical spectroscopy over wide fields is needed for transient classification and characterization. Specifically, a massively multiplexed spectrograph with rapid repointing capabilities (able to promptly slew to the location of the rarest and most interesting transients while also acquiring large samples of spectra of known classes of transients) would be uniquely

positioned to maximize the scientific return from time-domain astronomy in the next decade. Radio facilities with good (~μJy) flux sensitivity, sub-mas angular resolution, and 1–100 GHz coverage are needed to map the kinetic energy structure of the explosion ejecta, look for the presence of relativistic jets, and constrain the onset of pulsar wind nebula-like emission. Broadband X-ray capabilities with high sensitivity (10^{-19} erg/s/cm^2 @ < 10 keV; 10^{-14} erg/s/cm^2 @ 10–100 keV) are needed to map the kinetic energy structure of the ejecta, uncover the presence of relativistic jets, and map the media surrounding SNe, tidal disruption events, and NS mergers. Deep ultraviolet/optical/near-infrared spectroscopy is needed to constrain the chemical composition of the ejecta, with particular emphasis on nebular phase spectroscopy and spectroscopy of distant kilonovae discovered by GW detectors. Except for the very nearest events, ELTs will be required for late-time kilonova spectroscopy. An additional crucial need is for co-observing and rapid response capabilities across all observatories.

B-Q3. WHY DO SOME COMPACT OBJECTS EJECT MATERIAL IN NEARLY LIGHT-SPEED JETS, AND WHAT IS THAT MATERIAL MADE OF?

Relativistic jets—collimated beams of ejected material moving at nearly light-speed—are observed in a variety of systems: SMBHs in AGN, stellar-mass BHs and NSs in X-ray binaries, GRBs, and tidal disruption events (TDEs). Although jets have been intensively studied for many years with a variety of observational and theoretical techniques, they are still poorly understood. Several recent developments make this area ripe for progress. The Event Horizon Telescope (EHT) imaged the base of the jet in the nearby AGN M87 at an angular resolution comparable to the projected Schwarzschild radius. The panchromatic detection (radio to gamma ray) of an off-axis relativistic jet in the NS-NS merger event GW170817 discovered by LIGO/Virgo enabled the first constraints on jet structure in a transient (how energy is distributed within the jet). Atmospheric Cherenkov telescopes discovered TeV gamma rays from GRB jets. IceCube discovered extragalactic TeV-PeV neutrinos, which plausibly originate in powerful relativistic jets. Numerical simulation codes are now capable of first-principles investigation of magnetohydrodynamical jets and particle acceleration. Computer hardware (CPU/GPU) is approaching the power required to simulate problems in three dimensions. All of these nascent developments are poised for explosive growth.

B-Q3a. How Do Jets Launch and Accelerate?

At its base, a relativistic jet may be initiated and powered by rotating gas in the accretion flow via magnetic fields, gas pressure, or radiation pressure. Alternatively, jets may draw power directly from the spin energy of the BH via frame-dragged magnetic fields. Presently, there is no robust observational evidence favoring any proposed jet-launching mechanisms. If we could find evidence that relativistic jets extract power directly from spinning BHs through some combination of theory, simulation, and ultra-high-resolution observations of jet-launching regions, it would be a spectacular demonstration that BHs do not just consume mass and energy—that they sometimes also return energy to the world outside.

Accreting NSs sometimes produce relativistic jets (usually during a hard-to-soft spectral state transition), and there are indications that accreting WDs can also have jet-like activity. How similar are the observational properties of jets from different types of compact objects? This could provide clues to the jet launching mechanism. Can we numerically simulate the launching of NS jets? Relativistic jets in both long- and short-duration GRBs may be launched from either BHs or highly magnetized NSs (magnetars). Can we combine theory and observations to discriminate between these possibilities? Do WD jets move relativistically—and if so, what physical process drives such rapid ejection from such shallow gravitational potentials?

The Lorentz factor of the jet in M87 has been mapped as a function of distance from the central SMBH. The acceleration appears to be gradual, and the peak Lorentz factor is reached in only ~10^6

Schwarzschild radii. Is this behavior universal? What is the physics that controls the rate of acceleration? What determines the maximum Lorentz factor of the jet? What role do the available power, the amount of baryon loading, and drag from an external medium (ISM, stellar envelope, AGN cocoon) play? Why do stellar-mass BH jets reach Lorentz factors of only a few, while AGN jets (especially in luminous blazars) reach up to a few tens, and GRB jets reach up to a few hundreds?

B-Q3b. What Are Jets Composed Of and How Are Particles Accelerated Within Them?

Leptons (electrons/positrons), baryons (protons/nuclei), and magnetic fields are all believed to play a role in determining jet dynamics. But the relative importance of these components is not known. Is the jet launched as an electron-positron pair plasma or as an electron-ion plasma? How much baryon contamination does the jet subsequently experience, and where does most of it happen? What fraction of the power is carried by the magnetic field, and how does it vary with distance from the central object? The fraction of EM emission in jets that is hadronic versus leptonic in origin has important implications for energy requirements (by several orders of magnitude) and for AGN feedback.

Relativistic jets produce nonthermal EM radiation, indicating the presence of electrons and positrons with power-law energy distributions and emitting synchrotron and inverse-Compton radiation. How are these particles accelerated, and what determines their energy distribution (minimum and maximum Lorentz factor, slope of the energy distribution)? Theory suggests several candidate acceleration mechanisms: Fermi acceleration in shocks, magnetic reconnection, and shear acceleration. Which of these dominates in any particular system? Could multiple mechanisms operate in the same system, either co-spatially or at different locations along the jet? The particle acceleration processes studied in jets are broadly relevant to many other areas (solar and planetary physics, space physics, plasma astrophysics, etc.) as well.

B-Q3c. Are TeV-PeV Neutrinos and Ultra-High-Energy Cosmic Rays Produced in Relativistic Jets?

Ultra-high-energy cosmic rays (UHECRs, in the ~EeV range) can contribute significantly to the reionization of the IGM at the epoch of formation of the first AGNs and GRBs. They are also an important pressure component in the IGM of galaxy clusters as they form and virialize. UHECRs are widely surmised to be accelerated by shocks in the relativistic jets of AGNs or GRBs, with other sources also possibly coming into play. Can we verify that this idea is correct? Direct identification is difficult, because UHECRs are charged particles and thus lose directionality as they diffuse through magnetic fields. However, a clear directional signature would be provided by the high-energy neutrinos that UHECRs produce via interaction with the ambient photons in jets. Is this the origin of astrophysical high-energy neutrinos? An isotropic TeV/PeV neutrino background of astrophysical origin has been identified, but its origin in specific sources remains uncertain. Attempts at positional and temporal correlations with EM-detected bright GRBs have so far yielded negative results, but a 3σ positional and temporal correlation between gamma-ray flares and a high-energy neutrino associated with the blazar TXS 0506+056 is suggestive. Stacking analyses on other similar AGNs, however, indicate that additional types of sources may also need to be considered in order to account for the entire TeV-PeV neutrino background. Increased angular resolution and sensitivity of neutrino detectors are needed to address this.

Hadronic interactions between UHECRs and photons produce a comparable amount of energy in secondary gamma rays and neutrinos in the GeV-TeV range or above. Most of the higher energy gamma rays will cascade down to lower energies via γ-γ interactions, so that some of the observed GeV or lower-energy photons may be owing to this. In AGNs, would such hadronic cascades fill in the saddle point between the high- and low-energy humps of the EM spectral energy distribution? Does the lack of such radiation rule out significant UHECR acceleration in jets, or does it instead imply a low photo-hadronic optical depth on the target photons? These are questions for both theorists and observers.

Relevant Measurements and Capabilities

High-angular-resolution radio/millimeter imaging will address critical questions of jet launching and acceleration. With better (sub-mas to μas) angular resolution, polarization data for monitoring the magnetic field, and multiple observations to explore dynamics, one will be able to answer many key questions on jet launching. Improving angular resolution to a few μas and sensitivity to 1 mJy would increase the number of SMBH targets where the base of the jet can be imaged beyond the two or three currently accessible. Studying structure and acceleration along the length of a relativistic jet calls for high-angular-resolution imaging in several bands. Sub-milliarcsecond angular resolution at cm wavelengths, with full polarization, would provide information on the jet inclination, lateral structure, magnetic field strength, and variation of Lorentz factor with distance for many objects. High-angular-resolution imaging in the optical and infrared with <10 mas resolution and in X rays with sub-arcsec resolution would provide substantial new empirical information on the jet acceleration zone. Significantly improved capabilities for radio, optical, X-ray, and gamma-ray polarimetry will be invaluable for exploring the magnetic fields and composition of jets, as well as the presence of shocks and the field orientation with respect to the shock normal, which impacts the particle acceleration efficiency. High-angular-resolution Faraday rotation measurement in radio will be particularly critical for these goals.

General-relativistic magnetohydrodynamics (GR-MHD) codes are now able to follow jets from launch out to many orders of magnitude in distance, precisely the region where the above instrumental capabilities will provide observational data. Codes running on fast GPUs are seeing large speed increases over CPU-based codes. Training young scientists in the use of these specialized techniques, as well as investments in the relevant hardware, will be critical in this growth area.

High-frequency (Hz/kHz) GW observations of NS mergers and low/medium-energy gamma-ray follow-up will be able to measure the time between initial energy release and the first EM emission from jets in short-duration GRBs. Combining the information on jet inclination angle from GW and EM observations is required to explore the nature and evolution of the jet structure.

The energy distribution of nonthermal electrons in a jet can be deduced from observations of the broadband synchrotron spectrum. Spatially resolved spectra will show where particles are accelerated and how their energy distributions evolve, providing key constraints for theoretical models. The particle acceleration process itself, whether by shocks, reconnection, or shear, is best studied via numerical particle-in-cell (PIC) simulations, with 3D simulations about to become more routine. We can expect to understand the slope of the particle energy distribution for each acceleration mechanism, and the nature of the lower cutoff in the particle Lorentz factor. Determining the maximum Lorentz factor is more challenging. Support for theoretical and computational work and investment in computer hardware are essential for progress in this field.

For high-energy neutrinos, an effective detector volume an order of magnitude larger than IceCube and increased support for data analysis capabilities are required for detecting and localizing neutrinos from individual sources like TXS 0506+056, and to significantly increase the neutrino sample to enable studies of clustering and spectra. For rapid follow-up and correlation studies, continued support of *Swift* and *Fermi* spacecraft operations will be crucial until newer missions replace them. Also important are observations to study UHECR clusterings, anisotropies, composition, and spectra. Last, wide-field MeV-GeV gamma-ray facilities will be essential for identifying counterparts to high-energy neutrino and UHECR sources, along with wide-field and follow-up capabilities in the GeV-TeV band to extend these observations to higher energies for nearby sources.

B-Q4. WHAT SEEDS SUPERMASSIVE BLACK HOLES AND HOW DO THEY GROW?

While it is well established that SMBHs and galaxies grow together over cosmic time, the physics of both what seeded SMBHs in the first place and the processes that govern their growth remain poorly understood. This issue has couplings across astrophysics. Questions of accretion physics connect to AGN

phenomenology and feedback as well as to X-ray binaries and other accreting stellar-mass compact objects. The nature of BH seeds—and the rate and timing of their growth—is also an important factor in understanding the sources of cosmic reionization.

There are unprecedented opportunities ahead for determining the nature of the seeds of SMBHs and understanding the physics of how they have grown to the present-day population. Gravitational wave observatories like LISA and pulsar timing arrays will detect merging BHs over the 10^4–10^{10} $M_{\odot}$ range. Observatories are proposed with high sensitivity and high angular resolution across the EM spectrum, offering dramatic leaps in our understanding of the IMBH/SMBH population and growth mechanisms. Currently, our knowledge of super-Eddington accretion is rapidly growing on both the theoretical and observational fronts with the discovery of high-redshift massive quasars and NS ultra-luminous X-ray sources, as well as the first global 3D radiation magnetohydrodynamics simulations with realistic radiative transfer. In the next decade, a new era of 3D models of BH accretion—fully accounting for general relativity, radiation, and magnetohydrodynamics—will extend our understanding to the radiatively efficient accretion regime and provide physical calculations of radiative efficiency, accretion efficiency, and the launching of outflows and jets.

B-Q4a. How Are the Seeds of Supermassive Black Holes Formed?

At high redshift, SMBHs may originate with light seeds (~10^2 $M_{\odot}$; the compact remnants of the first generation of stars), intermediate-mass seeds (~10^3–10^4 $M_{\odot}$; from gravitational runaway in dense star clusters), or heavy seeds (~10^4–10^6 $M_{\odot}$; from direct collapse of gas in high-redshift halos). We will soon have the potential to discriminate between these models and determine the primary channel of SMBH seeding. One important diagnostic is the mass distribution of BHs at high redshift. Will we find evidence for a very large population of ~100 $M_{\odot}$ BHs at $z \gtrsim 10$, or will observations imply a smaller number of ~10^5 $M_{\odot}$ BHs? Another powerful test of seed models is the population of IMBHs (~10^2–10^4 $M_{\odot}$) in local galaxies. How many off-center IMBHs are there, and what is their mass distribution? If SMBHs were formed from light seeds, there should be many IMBHs that failed to merge into a galaxy's central SMBH and survive today as wandering BHs.

B-Q4b. How Do Central Black Holes Grow?

In order to understand what seeded SMBHs at high redshift, it is important to also understand the physics of SMBH growth. Because we can observe BHs only when they are growing, we need to understand which BHs are growing and why in order to extrapolate to the broader population. Furthermore, if the seeds of SMBHs are light, then BH growth is by necessity more efficient at high redshift, so the rates and efficiency at which BHs gain mass provide an additional test of seed models.

In the coming decades, important progress can be made in understanding the role of BH-BH mergers in forming the population of SMBHs seen today. What is the rate of SMBH mergers as a function of mass (~10^2–10^{10} $M_{\odot}$) and redshift (out to z ~ 20)? What fraction of binary (bound) and dual (neighboring but unbound) SMBHs merge? By answering these questions and measuring the spin distribution of IMBHs and SMBHs, the role of mergers in growing the SMBH population can be determined.

We are also poised to make strides in understanding BH accretion, which is also essential for understanding the physics of BH growth. Is super-Eddington accretion important in growing SMBH seeds? New sensitive facilities will be able to measure the accretion signatures of ~10^5 $M_{\odot}$ BHs at z ~ 10. Meanwhile, theoretical work and observational studies in the more local universe will shed light on the physics of super-Eddington accretion and its EM signatures. Lower accretion rates are also important for SMBH growth, so we must understand the structure and stability of sub-Eddington quasar accretion disks

as well as their radiative and accretion efficiencies. Last, a better understanding of the population of tidal disruption events will determine how often SMBHs swallow stars, and whether stellar tidal disruption is a significant contributor to SMBH growth.

Relevant Measurements and Capabilities

A range of planned facilities will be instrumental in enabling this science. *LISA* will detect mHz GWs from mergers of 10^4–10^7 $M_\odot$ BHs out to $z \sim 20$, yielding measurements of SMBH mass, spin, and merger rate. It will also reveal IMBHs through extreme mass-ratio inspirals. In the coming decade, pulsar timing arrays will detect a background of nHz GWs from the population of more massive ($\gtrsim 10^9$ $M_\odot$) SMBH binaries and mergers throughout the universe, and potentially individual SMBH binaries out to $z \sim 1$. High-frequency (Hz-kHz) ground-based GW observatories will detect mergers of BHs straddling the stellar-mass to intermediate-mass divide (~10–10^3 $M_\odot$). Currently planned detectors will reach to $z \sim 1$, while more advanced detectors could reach to $z \sim 20$. With this full-spectrum GW coverage, we will understand the population of merging BHs more than 10 orders of magnitude in mass. Planned time-domain optical surveys (including Rubin/LSST and other less-sensitive but higher-cadence facilities) will probe the population of tidal disruption events and find binary AGNs through periodicity searches. WFIRST will constrain the presence of IMBHs in the Local Group by discovering hyper-velocity stars. JWST will be able to efficiently establish a sample of high-redshift AGNs.

Further in the future, this science requires sensitive (~μJy), high-angular-resolution (sub-mas) radio imaging for finding accreting IMBHs and imaging binary AGNs. Sensitive (~10^{-19} erg/s/cm^2) X-ray observations with sub-arcsec angular resolution will also find accreting IMBHs and will additionally enable imaging of more widely separated dual AGN, measurements of SMBH spins, and—in concert with JWST spectra—measurement of the luminosity function of high-redshift AGN. Last, enhanced support for theoretical efforts is needed to enable accretion simulations and models of seed formation.

DISCOVERY AREA: TRANSFORMING OUR VIEW OF THE UNIVERSE BY COMBINING NEW INFORMATION FROM LIGHT, PARTICLES, AND GRAVITATIONAL WAVES

Astrophysical observations with non-EM messengers such as GWs, neutrinos, and UHECRs provide a new way to view the universe. Multimessenger astrophysics, where these new observations are combined with more traditional data across the EM spectrum, opens enormous discovery space for understanding high-energy astrophysical sources, and provides new cosmological tools and tests of fundamental physics. For decades, there were only two examples of source-specific multimessenger detections, both in MeV neutrinos: from the solar interior (starting in the 1960s), and from the nearby core-collapse supernova SN 1987A.

In the past decade, however, multimessenger astrophysics has come of age. We have seen the advent of GW astronomy and the first detection of GWs and photons from the same astrophysical source. We have also seen the discovery of astrophysical high-energy neutrinos and a potential association with a specific astrophysical source. Last, we have obtained new constraints on cosmic rays within the Milky Way and extragalactic cosmic rays, including more precise measures of the spectrum and composition of UHECRs, the discovery of TeV halos, and the discovery of pevatrons (PeV cosmic ray sources) in our galaxy. There is enormous potential in multimessenger astrophysics in the next decades, driven by improvements in ground-based GW detectors and neutrino observatories, by the advent of space-based GW observatories, and the maturation of pulsar timing arrays. Multimessenger astrophysics with these new messengers will be enabled by wide-field and rapid-response facilities across the EM spectrum for identification of EM counterparts and detailed follow-up studies. A few examples of this potential are given below.

B-DA1. Compact Binary (NS-NS and BH-NS) Mergers

Coordinated observation of compact binary mergers in both GWs and EM radiation is still in its infancy. However, improvements in ground-based GW interferometers combined with aggressive EM follow-up with existing and future facilities will usher in an era of population studies of NS mergers. As shown by our experience with GW170817, the combination of near-simultaneous gamma-ray and GW detections with rapid optical/infrared follow-up and gamma-ray/X-ray/radio monitoring can yield critical information. The next generation of instruments will enable detailed mapping between initial merger conditions (as determined by GWs and gamma-ray emission onset) and the merger outcome (e.g., BH or massive NS, jet/no-jet, jet physics/profiles), ejecta mass, ejecta chemical composition and *r*-process nucleosynthesis, and environment, as determined by kilonovae and long-term afterglows. We can hope for even more information from future nearby and favorably oriented events. In particular, had GW170817 been observed on-axis, the predicted TeV-PeV neutrino flux might have been detectable by current facilities, providing a new direct probe of particle acceleration and jet conditions for the first time. The detection of even a single NS merger in both high-energy gamma-rays and neutrinos would provide unprecedented data, emphasizing the importance of having sufficiently sensitive instruments to enable this. The combination of GW and electromagnetic measurements will also provide standard siren probes of cosmology.

B-DA2. Astrophysical TeV-PeV Neutrino Sources

Neutrino astronomy has begun. Ongoing observations and improvements in TeV-PeV neutrino observatories will yield a large sample of astrophysical neutrinos over the next decade. This may enable identification of EM counterparts of neutrino sources, thus clarifying their origin. Are there multiple neutrino source classes? For known source classes, what characteristics determine neutrino intensity? Another exciting possibility would be the discovery of positional coincidences between the highest-energy UHECRs and high-energy neutrinos. The prospects for such an identification will be greatly enhanced by future space-based UHECR observatories. Last, spatial, spectral, and composition measurements of UHECRs, combined with measurements of gamma-ray and both high-energy (TeV-PeV) and ultra-high-energy (EeV) neutrino diffuse emission, will establish the relationship between these quantities and possibly lead to a unified model to explain their origin.

B-DA3. Binary SMBHs

Groundbreaking near-future observations will be provided by low-frequency GWs with pulsar timing arrays (PTAs) and LISA. PTAs will likely detect the nHz GW stochastic background from the ensemble of $>10^8\ M_\odot$ SMBH binaries in the universe within the next few years, providing information about how SMBHs grow and evolve. By the end of the decade, PTAs could resolve multiple individual sources from either the closest SMBH binaries or the most massive ($>10^9\ M_\odot$). Those individual GW sources have the potential for exquisite EM follow-up, as the years-to-decades long periodicities of the SMBH binaries may manifest as variability across the EM spectrum. All-sky EM surveys may even facilitate PTA GW detections by initially identifying the most compact and nearby SMBH binaries, dramatically decreasing the size of the usual blind-search GW parameter space.

On a somewhat longer time scale, LISA measurements of mHz GWs will detect essentially every 10^4–$10^8\ M_\odot$ SMBH merger in the universe, providing fundamental information on SMBH evolution. These sources may be accompanied by counterparts covering a broad range of the EM spectrum and a wide span of time scales, including transients before, during, or after the merger as well as persistent counterparts. LISA localizations will be ~10 arcmin^2, offering the opportunity to find these EM

counterparts and enable multimessenger studies of sources and their host galaxies. Combining LISA data with the highly complementary measurements provided by photons and particles will enable transformative multimessenger science, including the birth and growth of supermassive BHs, and the expansion rate of the universe through standard siren measurements.

B-DA4. Galactic Ultracompact WD Binaries

LISA will detect thousands of ultracompact WD-WD binaries (and potentially many WD-NS binaries) in our galaxy that may be otherwise unidentifiable as such, or undetectable. Combining LISA detections with optical photometric, spectroscopic, and astrometric surveys (and with X-ray surveys for WD-NS binaries) will allow for a much more complete census of all types of WD binaries in our galaxy, including their mass distribution. It will also provide crucial constraints on progenitor models for type-Ia SNe and other transients and merger products. The WD-WD systems already identified as strong LISA candidates through optical studies show that this will be a powerful multimessenger combination.

B-DA5. Diffuse Thermal Background from Core-Collapse Supernovae

In this decade, Super-Kamiokande (with gadolinium loading added for improved sensitivity) may provide the first detection of the diffuse thermal neutrino background expected from the cosmic history of core-collapse SNe. The inclusion of Gd loading in its successor, Hyper-Kamiokande (an MeV neutrino experiment currently under construction in Japan), will provide significantly better sensitivity to this background with tens of events expected per year, ushering in a new era in neutrino astronomy. Combined with EM surveys of star formation and SNe, detection of the MeV neutrino background will provide a multimessenger connection to other tracers of core-collapse supernovae throughout the universe. The measured flux and spectrum of the neutrino background will provide information on the fraction of optically dark/unseen SNe, the fraction of core collapses that produce BHs, and important integrated constraints on the cosmic star formation history and the chemical enrichment of the universe from massive-star and SN nucleosynthesis.

B-DA6. A Supernova Within Our Own Galaxy

The ultimate multimessenger event would be a core-collapse SN within our galaxy. This event would produce a large flux of neutrinos, nuclear MeV gamma-ray line emission, and broadband emission across the EM spectrum, and perhaps high-frequency GWs as well. Note that, conservatively assuming a supernova rate within our galaxy of only one per century (most published estimates are higher than this), the Poisson probability of at least one Milky Way massive star supernova occurring in the next 20 years is 18 percent. Given the transformational science return expected, it is worth planning seriously for this possibility.

MeV neutrino (and possibly GW detectors) would see such an event first, with many thousands of neutrinos detected over a few seconds in current and near-future detectors. These neutrino and GW detections would provide early warning and degree-scale localization for the full suite of humanity's follow-up facilities to be deployed hours before shock breakout from a red supergiant progenitor (or perhaps just minutes before for compact stripped-envelope stellar progenitors). As in GW follow-up for NS-NS mergers, very wide-field monitors are necessary for quick identification, particularly in the infrared given the large optical extinction along lines of sight in the galactic plane. Direct diagnostics of the explosion mechanism and the properties of the neutron star in formation (e.g., rotation, convection) could be gleaned from simultaneous neutrino and GW detections in the first seconds after collapse. Neutrino flavor information would inform neutrino physics and our understanding of SN nucleosynthesis.

If it occurred, the transition from NS to BH would be imprinted on these signals, with profound implications for our understanding of this process. The late-time neutrino emission would directly inform our incomplete understanding of the birth of NSs. As evidenced by the famous few previous Milky Way core-collapse SNe (the Crab, Cas A, and SN 1987A), such an event would be studied for centuries.

The Kepler and Tycho SN remnants demonstrate the need to be equally prepared for the next Milky Way type Ia SN. Indeed, the expected rate is of the same order as that for core-collapse SNe. Here, the primary overlap in messengers is between the potential LISA detection of mHz GWs from the compact WD binary progenitor before explosion and information from the EM regime: MeV gamma-ray line emission from ejected nuclear products, multiwavelength continuum radiation from the SN including nonthermal emission from shock acceleration, and possibly a direct connection to the EM progenitor, which may be identified in existing catalogues. Without neutrino or high-frequency GW triggers to provide real-time advance information, it will be LISA, ultra-wide-field all-sky EM monitors, amateur astronomers, and our own eyes that may alert us to an event as it begins.

B-DA7. Fundamental Physics

The science return from multimessenger astrophysics extends beyond astrophysics. The extreme energies involved in interactions of UHECRs (and production of associated neutrinos and EM radiation) explores particle interaction cross sections at energies well beyond those achievable in terrestrial particle accelerators, allowing the study of exotic particle physics models. The relative arrival times of GWs and gamma rays from NS mergers provides the best measurement of the speed of gravity and a test of the weak equivalence principle. The relative arrival times of neutrinos and gamma rays from the same astrophysical event would provide a complementary test of the weak equivalence principle. NS formation, NS-NS mergers, and NS mass and radius measurements all constrain the equation of state of ultradense matter and the phase diagram of quantum chromodynamics (QCD).

B-DA8. Other Possibilities

More speculatively, there is a wealth of additional phenomena that may emerge with multimessenger observations in the next two decades. EM emission from stellar-mass BH mergers may be seen. With improvements in the sensitivity of ground-based GW interferometers, continuous GWs from a known radio or X-ray pulsar may be detected, providing the first measurement of a NS quadrupole moment. We may find new sources of both high-energy and thermal neutrinos from sources such as SN shock interactions, long-duration GRBs, tidal disruption events, and classical novae. We may be able to use heavy cosmic-ray abundance measurements in the Milky Way to constrain sites of *r*-process nucleosynthesis. Last, with ongoing and improved GW observations at all frequencies, we may find anomalous GW events that challenge general relativity, such as violation of the no-hair theorem, non-GR ringdown, or entirely new and unanticipated classes of events. With current and upcoming facilities for multimessenger astrophysics, we are opening a vast new discovery space. This virtually ensures that the most exciting new results will be in entirely unexpected areas.

Relevant Measurements and Capabilities

The key requirement to maximize the science return in multimessenger astrophysics is a broad range of facilities operating contemporaneously. The specifics of the needed capabilities for the individual messengers and EM bands are discussed in previous sections and are summarized in Table B-2.

TABLE B.1 Key Science Questions and Discovery Area

Question	Subquestions
B-Q1: What are the mass and spin distributions of neutron stars and stellar-mass black holes?	**B-Q1a:** What do the mass and spin distributions tell us about neutron star and black hole formation and evolution? **B-Q1b**: What is the population of noninteracting or isolated neutron stars and stellar-mass black holes? **B-Q1c:** What is the equation of state of ultradense matter?
B-Q2: What powers the diversity of explosive phenomena across the electromagnetic spectrum?	**B-Q2a:** When and how are transients powered by neutron stars or black holes? **B-Q2b:** When and how are transients powered by shocks? **B-Q2c:** When and how are transients powered by radioactivity? **B-Q2d:** What are the unexplored frontiers in transient phenomena?
B-Q3: Why do some compact objects eject material in nearly light-speed jets, and what is that material made of?	**B-Q3a:** How do jets launch and accelerate? **B-Q3b**: What are jets composed of and how are particles accelerated within them? **B-Q3c**: Are TeV-PeV neutrinos and ultra-high-energy cosmic rays produced in relativistic jets?
B-Q4: What seeds supermassive black holes and how do they grow?	**B-Q4a:** How are the seeds of supermassive black holes formed? **B-Q4b**: How do central black holes grow?
B-DA: Transforming our view of the universe by combining new information from light, particles, and gravitational waves	

TABLE B.2 Required Capabilities

Capability	**Science Enabled**	**Current/ Expected Facilities**	**Future Needs**
Radio time-domain surveys	B-Q1: ms-PSR searches and timing B-Q2/DA: FRB searches; transient detection	GBT, Arecibo, CHIME, FAST, JVLA, SKA, and Pathfinders	Multipixel cameras for single-dish pulsar observations. Pulsar search/timing backends for arrays. Arcsec localization for transients. Commensal FRB searches for all cm-band observations.
High-angular-resolution radio/mm imaging and polarimetry	B-Q2/DA: Transient follow-up B-Q3: Jet formation, acceleration and composition; particle acceleration B-Q4: Accreting IMBHs; binary AGN	JVLA, MeerKAT, GMRT, ATCA, ALMA, VLBI, EHT, SKA	Extremely high angular resolution (sub-mas to μas). Polarimetry and Faraday rotation.
O/IR time-domain surveys	B-Q1: Noninteracting binary or free-floating NSs and BHs B-Q2/DA: Transient detection; pre-explosion imaging of SNe B-Q4: TDEs in IMBHs; binary AGN	ASAS-SN, ZTF, Rubin/LSST, APOGEE, DESI, SDSS-V, ATLAS, Gaia, TESS, WFIRST, Euclid	Broad range of cadences (hours to weeks) and sensitivities (magnitude 10–24 in single images); prompt public release.
Massively multiplexed O/IR spectroscopy	B-Q1: Noninteracting binary NSs and BHs B-Q2/DA: Transient follow-up	APOGEE, DESI, SDSS-V	Rapid response (<1 hr). Cadences of hours to weeks. ELT-class sensitivity. $R \sim 1000$.
Deep O/IR line spectroscopy	B-Q1: Radial velocity curves of binaries of interest B-Q2/DA: Transient follow-up B-Q4: redshift of high-z AGN	8–10 m-class ground, HST, JWST	ELT-class sensitivity. Rapid response to transients. $R \sim 100$ for classification, $R \sim 1000$–5000 for follow-up and RVs.
High-angular-resolution O/IR imaging and spectroscopy	B-Q3: Jet acceleration; particle acceleration B-Q4: Dynamical confirmation of local IMBHs; binary SMBHs	8–10 m-class ground AO, HST, JWST, WFIRST	<10 mas angular resolution and ELT-class sensitivity. Rapid response to transients. $R > 5000$ for IMBH masses.
UV imaging and spectroscopy	B-Q2/DA: Transient follow-up	Swift/UVOT, HST	Comparable post-Swift and post-HST coverage. Rapid response to transients.
Wide-field X-ray (0.5–100 keV) monitors	B-Q1: New NS/BH transients B-Q2/DA: Transient detection	Swift/BAT, MAXI, Fermi/GBM, eROSITA	Post-Swift and post-Fermi coverage. Range of capabilities optimizing trades between high sensitivity, wide-field coverage, and <arcmin localization.
X-ray imaging and spectroscopy	B-Q1: NS/BH disk reflection lines B-Q2/DA: Transient follow-up B-Q3: Jet spectroscopy	Chandra, XMM, NICER, NuSTAR, XRISM, Athena	10^{-19} erg/cm^2/s sensitivity and moderate ($R \sim 100$) spectral resolution. Hard X-ray coverage (10–100 keV) with 10^{-14} erg/cm^2/s sensitivity. Rapid response to transients.
X-ray spectral timing	B-Q1: NS/EOS pulse profile modeling; pulsar timing B-Q2/DA: Transient follow-up	XMM, NICER	Post-NICER/XMM coverage. <0.1 ms time resolution. Larger effective area (>1 m^2 @ 1 keV; >4 m^2 @ 10 keV). High throughput.

Capability	Science Enabled	Current/ Expected Facilities	Future Needs
High-angular resolution X-ray imaging	B-Q1: ULXs and other point sources in nearby galaxies B-Q2/DA: Transient follow-up B-Q3: Jet acceleration, particle acceleration B-Q4: SMBH seeds	Chandra, XMM, NuSTAR, Athena	High angular resolution (<1 arcsec @1 keV; <15 arcsec @20 keV). Hard X-ray (>10 keV) coverage. 10× Chandra/NuSTAR sensitivity.
X-ray/gamma-ray polarimetry	B-Q3: Jet and disk orientation and geometry	INTEGRAL, IXPE	10× IXPE sensitivity. Soft X-ray and MeV gamma-ray coverage.
MeV gamma-ray line spectroscopy	B-Q2/DA: Nuclear lines from SNe	INTEGRAL/SPI	Wide (>1 sr) FOV. Sensitivity $<8 \times 10^{-6}$ ph/cm^2/s in 10^6 s for ~1 SN-Ia/yr detected.
MeV/GeV gamma-ray imaging	B-Q1: Faint ms pulsars B-Q2/Q3/DA: Transients; counterparts for neutrino/UHECR sources; GRB jet launch	Fermi	MeV coverage. Post-Fermi GeV coverage.
TeV gamma-rays	B-Q3/DA: Counterparts for neutrino/UHECR jet sources, EM sources; particle acceleration	HAWC, MAGIC, HESS, VERITAS, LHAASO	Post-HAWC/VERITAS coverage.
Low-frequency (nHz/mHz) gravitational waves	B-Q1: NS and BH binaries B-Q4: SMBH binaries B-DA: GW counterparts of EM sources	NANOGrav and other PTAs, LISA	Continued PTA coverage with larger pulsar sample. Detect all merging SMBHs, localize loudest to <10 arcmin^2. Full U.S. access to LISA data.
High-frequency (Hz/kHz) gravitational waves	B-Q1/Q3: NS and BH mergers/jets B-Q2: Transient detection B-Q4: IMBH mass function B-DA: GW counterparts of EM sources	LIGO/Virgo, KAGRA, LIGO-India, LIGO/A+	BNS mergers to $z \sim 10$; 30/30 $M_{\odot}$ BBH and IMBH mergers to $z \sim 20$. Localization to <10 deg^2.
MeV neutrinos	B-Q2/DA: SNe (including diffuse thermal background)	Super-K, Hyper-K	10× Hyper-K volume for ~1 SN/yr.
TeV/PeV/EeV neutrinos	B-Q3: Jet counterparts/composition B-DA: ν counterparts of EM sources; diffuse TeV/PeV background	IceCube, ANTARES, KM3NeT	10× IceCube volume for ~1 ν /yr from TXS 0506-like transients. EeV coverage.
Ultra-high-energy (EeV) cosmic rays	B-Q3: Jet counterparts/composition B-DA: UHECR counterparts of neutrino and EM sources	Auger, TA, LHAASO	Continued coverage. 10× larger exposure. >4× larger detector area.
Theory, computation, and simulations	B-Q1/Q2/Q3/Q4/DA	Broad support for theory and computation across all areas. Next-generation computing for multidimensional radiation hydrodynamics and PIC simulations, numerical relativity. Training for GPU-based computation. Advanced nuclear reaction network, cosmic ray transport, and hadronic cascade simulations.	

C

Report of the Panel on Cosmology

INTRODUCTION

We live in a remarkable time in human history when—for the first time—we can observe most of our universe and how it has evolved. The study of cosmology has been expanded far beyond Earth and the Milky Way to encompass vast extragalactic distances and the dramatic evolution of our universe. Enabled by the profound technological advances of the past century, the cosmology community has compiled exquisite measurements and made remarkable discoveries about the history and composition of the universe. The results have led us to a simple empirical cosmological model, referred to here as the standard cosmological model, that unifies a wide range of observational phenomena and provides a crisp starting point for astrophysical computations. This model has continued to successfully explain the measured evolution of our universe even as the body of data that might have challenged it has improved by orders of magnitude over the past two decades. Yet the standard cosmological model remains incomplete, lacking an underlying physical explanation of key ingredients. Realistic physical theories predict a wide range of observable signatures, and the opportunity to discover these signatures is the driving motivation for the coming decade of cosmological research.

The foundation of modern cosmology theory is the Hot Big Bang, in which an initially hot, dense, and nearly smooth universe rapidly expands and cools. Out of this early pressure cooker emerges the universe’s present-day composition: the familiar nuclei and electrons of normal matter, the relic heat now encased in the cosmic microwave background (CMB), a cosmic neutrino background, and an unknown dark matter that outweighs normal matter by a factor of six.

Using the well understood physics of plasmas, we are able to map the temperature fluctuations seen in the CMB back to the primordial conditions imprinted in the Big Bang. The small primordial fluctuations are inferred to closely follow a specific statistical pattern: Gaussian correlations with no preferred scale and with all components (i.e., dark matter, nuclei, photons, etc.) varying spatially together maintaining a fixed composition. While simple, this result is profoundly important because it indicates that the density perturbations were established *before* the Hot Big Bang phase of cosmic evolution. It is remarkable that these inferred properties match exceptionally well to the predictions of the theory of cosmological inflation, in which extraordinarily rapid expansion in the earliest moments of the universe established the large-scale homogeneity and flatness of the universe while also causing quantum fluctuations to create exactly the kind of density perturbations we observe.

As time passed, these primordial density perturbations grew in amplitude to form the detailed structure of the universe. Observations of this structure, in surveys of both galaxies and the CMB, clearly require something beyond normal matter to explain the experimental results. In the common paradigm, this is cold dark matter (CDM), some unseen gravitating material that moves nonrelativistically in the recent universe. The standard cosmological model posits the CDM as an empirical extreme—noninteracting, nondecaying, and without any thermal motion—but observations to characterize the properties of dark matter on galactic and sub-galactic scales are limited.

In addition, observations of the recent universe have shown that its expansion is presently accelerating. This remarkable discovery is not readily explained by a model containing only matter, but

instead indicates a new feature, dubbed dark energy. In the standard cosmological model, this is Einstein's cosmological constant: an energy and pressure characterizing empty space whose gravitational effect drives the acceleration. While next to nothing is known about the underlying cause of the acceleration, today's observations are consistent with the energy density being constant in time, as the cosmological constant would be.

There is no doubt that the standard cosmological model is a triumph. By adopting simple versions of inflation, dark matter, and dark energy, the model can match observational results despite orders of magnitude of improvement in cosmological measurements over the past 20 years. But there is also no doubt that the model is incomplete, as these essential components are not found within the standard model of particle physics. The panel stresses that the familiarity of the names of these components must not obscure this crucial problem. While Occam's razor favors the adoption of the simplest physical theory, the standard model of cosmology is not physically grounded, and particle physics models built to reproduce our cosmological observations almost invariably have observational signatures that deviate from the standard model—deviations we may well be able observe this decade.

STATE OF THE FIELD

In the Astro2010 decadal survey, the Panel on Cosmology and Fundamental Physics presented four questions—(1) How did the universe begin? (2) Why is the universe accelerating? (3) What is dark matter? (4) What are the properties of neutrinos?—as well as a discovery area in gravitational wave astronomy.[1] Progress on observational and experimental data sets to study these topics has been tremendous. Some notable highlights are:

- An explosion of arcminute-scale CMB data has extended the standard cosmological model to unprecedented precision, produced superb catalogues of galaxy clusters, and has opened the frontier of CMB lensing and the kinematic Sunyaev-Zel'dovich (SZ) effect.
- Searches for inflationary gravitational waves have improved in precision by more than a factor of 10, to the point where they disfavor many of the simplest models of inflation.
- Maps of large-scale structure have enabled a range of scientific advances, including (but not limited to) measurements of the cosmic distance scale over a wide range of redshift using the baryon acoustic oscillations (BAO).
- Cosmological weak lensing has leapt forward, with the uncertainty in the lensing-inferred amplitude of late-time density fluctuations decreasing four-fold to about 3 percent.
- Measurements of the Hubble constant from the direct distance scale and from strong gravitational lensing have improved to a precision of about 2 percent, while supernovae measurements at cosmological distances have driven precision on the dark energy equation of state w below 5 percent.
- The first gravitational wave events have been detected, including an initial application of the standard siren method of constraining the cosmic expansion rate. The observed travel time of gravitons resulted in an improvement of more than 10 orders of magnitude in the determination of the speed of propagation of gravitational waves.
- The primordial deuterium-to-hydrogen ratio has been measured to 1 percent precision.
- A wide range of searches for the astrophysical detection of dark matter have occurred, greatly improving the limits on many possible scenarios.

While much of the decade has been marked by a concordance between experimental results and the predictions of the standard cosmological model, not everything agrees. Direct measurements of the

[1] National Research Council, 2011, *Panel Reports—New Worlds, New Horizons in Astronomy and Astrophysics,* The National Academies Press, Washington, D.C., https://doi.org/10.17226/12982.

Hubble constant, H_0, tend to give higher values than those implied by CMB and large-scale structure data. The Experiment to Detect the Global EoR Signature (EDGES) project reports an 80 MHz spectral distortion consistent with 21 cm absorption from redshift 17, but with an amplitude several times larger than predicted. And there is a haze of gamma-ray emission that peaks at around 1 GeV from the inner Milky Way, consistent with a dark matter annihilation signal but also possibly explained as high-energy emissions from undetected pulsars. Whether resolving these discrepancies will ultimately require a new addition to the standard cosmological model is unknown, but they highlight the importance of a broad experimental program.

The coming decade will provide unprecedented cosmological opportunities. One of the major achievements of the past decade has been the development of a new generation of facilities, now under construction or in early operations, that will push the field dramatically forward. The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), Dark Energy Spectroscopic Instrument (DESI), Subaru/Prime Focus Spectrograph (PFS), Euclid, and the Nancy Grace Roman Space Telescope (formerly WFIRST) will provide superb optical and near-infrared imaging and spectroscopy surveys. The Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) will take observations further into the infrared with low-resolution spectroscopic mapping. The South Pole Telescope (SPT)-3G, Advanced Atacama Cosmology Telescope (ACT), and Simons Array will produce high-sensitivity maps of the CMB polarization at arcminute-scale resolution, while experiments such as the Keck Array, Background Imaging of Cosmic Extragalactic Polarization (BICEP)3, Spider, and Cosmology Large Angular Scale Surveyor (CLASS) will measure CMB polarization at large angles. The recently launched Extended Roentgen Survey with an Imaging Telescope Array (eROSITA) will produce a sensitive X-ray map of the full sky, with more detailed measurements of individual sources to be made by the Advanced Telescope for High Energy Astrophysics (ATHENA) mission early in the 2030s. Gaia will continue its mission, increasing its sensitivity to stellar proper motions and astrometric binaries. Gravitational wave (GW) observatories are rapidly extending their reach, including the ongoing development of the Laser Interferometer Space Antenna (LISA) mission. These surveys are complemented by a wide range of narrow-field facilities that allow us to pursue important companion and follow-on studies. The James Webb Space Telescope (JWST) is the coming exemplar of these facilities, but nearly every large telescope plays some role in cosmological science, spanning all wavebands from pulsar timing in the radio, to spectroscopy of faint transients in the optical, to dark matter annihilation searches in the gamma rays.

COSMOLOGY IN THE 2020S AND BEYOND

With both compelling mysteries and extensive observational means by which to explore them, this will be an amazing decade for cosmology. In this report, the panel identifies four major science questions for the upcoming decade: (1) What set the Hot Big Bang in motion? (2) What are the properties of dark matter and the dark sector? (3) What physics drives the cosmic expansion and large-scale evolution of the universe? (4) How will measurements of gravitational waves reshape our cosmological view? The panel also identified a discovery area: The Dark Ages as a cosmological probe.

These are familiar questions, but our experimental ability to tackle them is increasing rapidly and radically. As the panel explains below, the range of possible discoveries in these foundational areas is very broad. The motivation and capability to explore the unseen constituents and earliest moments of the universe remains one of the central themes of astrophysics.

It is also important to note the deep and increasing connection between cosmology, with its precision observations over enormous volumes, and the rest of astrophysics. Cosmological probes are invariably intertwined with their astrophysical context. These connections are often couched as "systematic uncertainties," which ignores the synergistic opportunities that come with the co-development of different areas of the field. An obvious example is galaxy formation, which is now highly tied to its

cosmological context. The need to embrace and extend these connections will only increase as the data become more sensitive and we seek more subtle cosmological signatures.

A major purpose of this report is to connect these scientific questions to the capabilities needed to pursue them. In the discussion of each question, the panel has identified the key observational and experimental facilities required. Where the panel has identified an important quantitative science goal, it has stated that goal. However, in many cases, discovering the physics behind our cosmology is an exploratory endeavor without a specific threshold. In these cases, the panel has opted to use the term “next generation” to indicate where future yet-unfunded facilities could provide important improvements, typically at the order of magnitude level. The practical selection of precise quantitative requirements will necessarily depend upon a balance of technical opportunities, cost, timing, and risk. The panel uses the term “funded upcoming facilities” to discuss the important role of facilities currently under fully funded construction; in some cases, the gains from these facilities will be so transformative that the panel believes it is critical to assess their early results before identifying opportunities for future generations of experiments.

C-Q1. WHAT SET THE HOT BIG BANG IN MOTION?

A vast number of observations allow us to characterize the state of the universe early in its history when it was hot, dense, and expanding rapidly. One of the fundamental discoveries of modern cosmology is that the primordial density fluctuations, the seeds of the structure of the universe observed throughout cosmic history, were created before the hot phase of the Big Bang. As a result, studying these primordial fluctuations provides a unique window into physics at extremely early times and at energy scales many orders of magnitude above what researchers can access in the laboratory.

The question of what process set the Hot Big Bang in motion and created the seeds of structure has been with us for many decades. Early theoretical developments, together with observations over the past two decades, have established the inflationary paradigm as the dominant picture in the field. In inflation, the universe went through an early period of accelerated expansion that smoothed out prior anisotropies, ending in a dramatic event that filled the universe with high-energy particles. The initial seeds for structure resulting from this period are expected to have simple properties: the statistics of the seeds follow almost perfect Gaussian correlations, invariant in scale, with no spatial variation in the composition among different particles. Deviations from these simple predictions carry most of the information about the inflationary period. In addition, inflationary theories always predict a stochastic background of gravitational waves whose amplitude and scale dependence, if measured, would provide important information about inflation as well as the quantum theory of gravity.

In the past decade, the Planck satellite measured departures from scale invariance of the power spectrum of density fluctuations, in line with the expectations of the simplest inflationary models, and placed exquisite constraints on departures from Gaussianity and fluctuations in the composition of the universe. The BICEP/Keck CMB polarization experiments put stringent upper limits on the amplitude of inflationary gravitational waves. We now stand at a crossroads for the inflationary paradigm because the improved measurements that could be performed in the coming decade will allow us to cross important theoretical thresholds and significantly improve our understanding of the inflationary epoch.

C-Q1a. Primordial Gravitational Waves

Gravitational waves are inevitably produced during an inflationary epoch and would survive to the present day. If inflation occurs at a sufficiently high energy scale, those gravitational waves can be observed through their imprint on the large-scale polarization pattern of CMB maps, the so-called B-modes. The past decade has seen a steady advance in orbital, sub-orbital, and ground-based experiments to measure CMB polarization at exquisite precision. The amplitude of this gravitational wave background

is quantified using the ratio of the amplitude of gravitational waves (tensor) to that of density fluctuations (scalar) produced during inflation, a ratio known as r. At the time of this report, the experimental constraints at $r < 0.06$ (95 percent confidence) already rule out very interesting portions of the parameter space of models.

A concerted effort over the next decade to improve the sensitivity to gravitational waves by a factor of 10–100 would cross important theoretical thresholds. In particular, a measurement of $r > 0.01$ would imply that the inflationary field moved over very large distances, larger than the Planck scale, in field space as inflation proceeded. Such an observation would be highly constraining to quantum gravity theories. Furthermore, models that naturally explain the scale-dependence in the density fluctuations observed by Planck by fixing the spectral index of density fluctuations to be inversely proportional to the number of e-folds of observable inflation predict $r > 0.001$. A next generation of large-angle high-sensitivity CMB polarization measurements along with arcminute-scale maps to provide the requisite control of foreground gravitational lensing can bring early universe cosmology to these two important discovery thresholds. Even if the gravitational waves are not detected, such limits would lead to a significant improvement in the understanding of the primordial universe.

C-Q1b. Non-Gaussianity of the Large-Scale Structure of the Universe

The statistical properties of the primordial density fluctuations in the universe encode information about the physical processes responsible for their generation. Minimal models of inflation involve a single field that evolves during inflation, serving as a clock that determines when inflation ends and the Hot Big Bang begins. Interactions between fluctuations of this clock field, or between such fluctuations and those of other fields during inflation, generically cause noticeable departures from Gaussian correlations in the distribution of the primordial structural seeds. If such departures can be detected in surveys of cosmological structure, then one can study these inflation-era interactions and constrain the physical origin of perturbations.

A wide range of possible non-Gaussian signals in the statistics of the primordial seeds are of cosmological interest, but one particular form contains an important quantitative threshold for next-generation surveys. In single-field inflation models, fluctuations correspond simply to time delays between different regions of space. This remarkable fact implies that a particular kind of deviation, called local-type non-Gaussianity, must be extremely small in these scenarios ($f_{NL,local} << 1$, in the usual parameterization). In contrast, if the observed density fluctuations originate from fields other than the inflationary clock field, or were not created during an inflationary period, interactions between the fields are generically not suppressed and produce large local-type non-Gaussianities in the primordial seeds, with $f_{NL,local}$ of order 1 or larger. As a detection of primordial local-type non-Gaussianity with $f_{NL,\,local}$ of order unity would falsify single-field inflation, the search for primordial non-Gaussianity, either to detect a signal or to constrain $f_{NL.\,local}$ to be below 1 with 5σ significance, is particularly important.

Advancing to this level of sensitivity will require three-dimensional surveys of very large volume, high sampling density, and exquisite large-scale systematic control to accurately measure a large number of fluctuation modes. The balance of volume, sampling density, and target properties could vary among viable surveys. NASA's SPHEREx mission, as well as redshift surveys such as from DESI and Euclid, will be the next step, but even larger surveys at higher redshift are needed to reach the target value. These will require next-generation high-multiplex spectroscopic facilities, likely in the optical and infrared but possibly at radio wavelengths, with the goal of mapping as many linear-regime modes of the primordial structural seeds as possible, at least a factor of 10 more than funded upcoming facilities. Cross-correlations of such surveys with CMB anisotropy maps can leverage the kinematic SZ effect to improve constraints on the largest-scale density perturbations.

C-Q1c. The Initial Power Spectrum of Density Fluctuations

In most scenarios for the production of primordial structural seeds, different spatial scales are generated at different times. This presents an opportunity to explore the history during the production era by measuring the amplitude of fluctuations over a wide range of scales. The physical processes governing the early universe could involve a number of additional degrees of freedom with a wide range of interactions, and in many cases, these dynamics reveal themselves in features on top of an otherwise smooth primordial power spectrum. Similarly, a detection of large-scale fluctuations between different species of matter would be an important new constraint on inflation and the thermal history of the universe. Detailed measurements of the primordial power spectrum can be advanced with new maps of the CMB, large-scale structure, and clustering of the intergalactic medium; in particular, the surveys required to measure non-Gaussianity will be excellent for measuring the large-scale power spectrum. Deviations at much smaller scales could be detected owing to their impact on CMB spectral distortions or on small-scale structure in early galaxy formation.

Summary of Capabilities Needed for C-Q1

Capabilities needed include a next generation of CMB polarization experiments (both large and small angular scales) to seek the primordial gravitational waves, and a next-generation large-volume redshift survey to seek primordial non-Gaussianity.

C-Q2. WHAT ARE THE PROPERTIES OF DARK MATTER AND THE DARK SECTOR?

Since the Astro2010 decadal survey, the field of dark matter theory and detection has undergone a paradigm shift. In previous decades, the field focused primarily on two dark matter candidates, weakly interacting massive particles (WIMPs) and axions, motivated mainly by their ability to solve long-standing open questions within the Standard Model of particle physics. However, recent work has emphasized that dark matter may arise from a dark sector more analogous to the visible sector of familiar particles, with its own dynamics and forces and with new terrestrial and astrophysical signatures. The nondetection of physics beyond the Standard Model at the Large Hadron Collider (LHC)—notably not finding signatures of supersymmetry—has served to further highlight the possibility that the dark sector need not be closely connected to well-recognized questions in particle physics. The breadth of new dark matter candidates and dark sector dynamics that have arisen from these recent explorations offers new motivation and opportunities to detect astrophysical signatures of dark matter.

C-Q2a. Dark Sector Signatures in Small-Scale Structure

The only irreducible interaction of the dark matter is through gravity, and it is through the gravitational interaction that all of our knowledge of dark matter in cosmology arises. As such, the way dark matter clusters gravitationally is a unique window into the nature of the dark matter and its attendant forces. Indeed, such measurements have already been very effective on super-galactic scales in the Milky Way for establishing the cold dark matter model. However, the clumpiness of dark matter on small scales is today only loosely constrained, save for the extreme case of objects compact enough to produce microlensing of stars. Many theories of dark matter beyond the WIMP paradigm feature modifications of the scale-invariant power spectrum—for example, resulting in gravitational collapse in the early universe into dark matter mini halos, which can be thousands or even a million times more dense than LambdaCDM sub-halos. New forces also generically appear in a broad range of dark matter models, giving rise to dark matter self-interactions and modifications of CDM predictions for the abundance and

density profile of halos. Such predictions and others provide compelling motivations to sharply extend our study of small-scale clustering.

The small-scale clustering of dark matter can be explored over a wide range of length and mass scales, ranging from the substructure of clusters, to dwarf galaxies, to sub-planetary-mass relics. This, in turn, relies on a large set of observational opportunities. Higher masses, above a million solar masses, can be sensitively probed by resolved gravitational lenses and by dwarf galaxy counts and mass profiles. These studies will be substantially advanced by the funded upcoming facilities LSST, Euclid, and Roman Space Telescope, and the panel notes the opportunity of the Atacama Large Millimeter/Submillimeter Array (ALMA) to characterize gravitational lenses. Next-generation large-aperture optical telescopes will be critical for spectroscopy of both dwarf galaxies and lens systems.

Stellar astrometric measurements, like those of galactic stellar streams and the survival or disruption of wide stellar binaries, probe intermediate masses from the dwarf-galaxy scale down to about a solar mass. This area is being revolutionized by Gaia, and augmented by both LSST and the funded upcoming wide-field high-multiplex optical/infrared spectrographs. At yet lower masses, possibly as low as 10^{-14} solar mass, pulsar timing arrays offer a novel opportunity by searching for timing anomalies owing to gravitational lensing by dark matter lumps passing between the observer and the target pulsar, even if the lumps are not compact enough to generate microlensing of the flux. Enhancing the network of radio telescopes capable of precision pulsar timing in order to substantially increase the precision, cadence, and sample size of timing measurements could considerably extend these low-mass searches.

C-Q2b. Dark Sector Imprints on Big Bang Nucleosynthesis and Recombination

The dark sector can leave other imprints on cosmic evolution. Current measurements still allow significant room for additional dark sector contributions to the early universe's energy density. One potential contribution—well motivated by numerous extensions of the Standard Model of particle physics—is from the relics of light particles produced thermally in the early universe, here called dark radiation. Dark radiation and its self-interactions can be constrained by arcminute-resolution CMB measurements of the recombination-era damping of waves in the baryon-photon fluid. With a next generation of such experiments, light-particle relics could be detected at an energy density only 1 percent to 2 percent of that of the cosmic neutrino background, allowing detection of relics that thermally decouple from the Standard Model before the quantum chromodynamics phase transition.

Measurements of the baryon density provide another window into the early universe. These come independently from CMB anisotropies and from the light-element abundances predicted from Big Bang Nucleosynthesis (BBN). These agree except for the ^{7}Li abundance, which shows a long-standing factor-of-two discrepancy. A measurement of the ^{7}Li abundance in low-metallicity diffuse gas could determine whether there is indeed an anomaly in the BBN predictions. If so, dark sector physics is a candidate explanation. For example, dark sector models can produce relativistic byproducts that lead to observable signatures in BBN. In addition, further improvements in the measurement of the helium and deuterium abundance can constrain theories of light dark matter. Such abundance measurements require high-resolution ultraviolet and optical spectroscopy with a next generation of larger aperture telescopes.

C-Q2c. Annihilation By-Products

While the LHC has not discovered supersymmetry or any signs of new physics at the weak scale, there are important models of supersymmetric dark matter that the LHC cannot reach, notably those in which the dark matter interacts only with the weak force and all the supersymmetric particles interacting directly with the strong force are too heavy to produce at the LHC. These models typically predict dark matter candidates at the few-TeV mass scale, with cross sections for annihilation to photons that make them observable with next-generation Cherenkov telescopes that have sufficient scope to detect an

annihilation cross section to two photons larger than 10^{-28} cm^3/s for dark matter masses of 1 TeV. This represents a crucial opportunity to search an otherwise-unreachable part of the weak-scale dark matter model space.

At the same time, at a much lower mass scale, the origin of the GeV gamma ray excess toward the galactic center remains unknown. It may be owing to dark matter or to millisecond pulsars. A goal of the next decade is to solve this puzzle. A next-generation gamma-ray telescope with better angular resolution than the Fermi satellite to resolve point sources in the galactic bulge, or better sensitivity to photons from dark matter annihilation in dwarf galaxies, would be a powerful tool for doing so. Less directly, deeper pulsar searches with next-generation radio telescopes could identify potential sources.

Summary of Capabilities Needed for C-Q2

The search for dark matter signatures is wide-ranging and exploratory, but the next generation of radio telescopes for pulsar timing, large-aperture optical telescopes, high-resolution CMB polarization mapping, GeV telescopes, and TeV-scale Cherenkov telescopes are particularly important to make progress in this field.

C-Q3. WHAT PHYSICS DRIVES THE COSMIC EXPANSION AND LARGE-SCALE EVOLUTION OF THE UNIVERSE?

One of the striking features of dark energy is that it explains not only the accelerating expansion seen in the late universe but also the cutoff of large-scale structure growth. Understanding the physics behind cosmic expansion requires testing both the expansion of space and the growth of structure across cosmic time. Together, these observations will provide an end-to-end test of our standard cosmological model and measure the properties and masses of neutrinos—the last known unweighed constituent of our universe. The panel stresses the importance of pursuing precise and accurate measurements that span a wide range of redshifts and clustering scales, and of doing so with methods that are complementary in both their systematic errors and their sensitivity to the physics of the cosmological model. Below, a number of key observations are highlighted to elucidate the physics behind cosmic expansion.

C-Q3a. The Physics of Cosmic Acceleration

One of the most profound discoveries in modern cosmology has been the accelerating expansion rate of the universe, which has now been independently confirmed by multiple probes. The cause of this acceleration, which has been dubbed “dark energy,” remains a mystery. The leading theory is that of a cosmological constant with a constant energy density over time, today comprising approximately 70 percent of the energy density of the universe. Testing for deviations from the cosmological constant model to the practical limits of available methods remains a key goal of the field of cosmology, as such deviations would be a signature of new physics or reveal a breakdown of general relativity (GR) at large scales.

It is critical to probe this acceleration with diverse and independent probes that provide constraints on both distance scales and the growth of structure. Key methods to explore this question in the coming decade include weak gravitational lensing of distant galaxies and the CMB, BAO and redshift-space distortion measurements from redshift surveys, supernovae Ia distance measurements, Hubble constant measurements, and galaxy cluster abundances. The funded upcoming wide-field survey facilities and CMB experiments will produce a major leap forward in both statistical reach and systematics control of these methods. Beyond currently funded capabilities, next-generation CMB surveys offer important new opportunities in both CMB lensing—an emerging field that will provide a

longer redshift lever arm for lensing studies of structure growth—and kinematic SZ studies, which combine CMB and redshift survey data to measure large-scale velocity flows. Beyond the major wide-field surveys, other capabilities will be needed to support these methods—for example, deep spectroscopic training data sets for photometric redshifts of weak lensing samples, follow-up telescope resources for Type Ia supernovae, and narrow-field instruments for local distance measurements and strong lensing cosmography. Narrow-field observations would be particularly advanced by the next generation of large-aperture optical telescopes.

The flexibility of the Roman Space Telescope mission will provide an important and powerful capability to investigate opportunities or questions raised by many of the funded upcoming wide-field survey facilities and CMB experiments, listed earlier in this appendix, that will begin earlier in the decade. The middle of the decade would be an excellent opportunity to assess progress and identify further scientific and technical opportunities in the field. The panel notes the particular importance of studying the low-redshift period where dark energy dominates the expansion rate. Opportunities for doing so include (1) standard candle and standard siren methods, which are not limited by the cosmic variance of large-scale structure, and (2) a next generation of densely sampled galaxy redshift and lensing surveys to test the impact of dark energy on structure formation with greater sensitivity.

In order to thoroughly test the predictions of the cosmological model, robust and model-independent tests for deviations from GR's prediction for the growth of large-scale structure are an essential element of the research program of the field. These tests largely use the same data sets as tests of cosmic acceleration. Existing data have motivated attempts to build modified gravity theories that explain this acceleration while remaining consistent with stringent constraints on gravity from other measurements—a challenging exercise that has deepened the understanding of GR and inspired new observational tests of gravity.

C-Q3b. The Properties of Neutrinos

The relic population of neutrinos forms an important dynamical component of the universe. This fact enables large-scale structure surveys to probe fundamental physics of neutrinos, such as their mass and possibly even their self-interaction cross section. Further, this characterization of neutrino properties will also be necessary to model the cosmological observables that feed into the dark energy measurements described above.

One key goal for next-generation surveys is to determine whether the neutrino mass hierarchy is "normal," with two similar lower-mass states much below the third, higher-mass state, or "inverted," with two similar higher-mass states well above the third, lower-mass state. Flavor oscillation experiments imply a minimum sum of the three masses of 0.06 eV in the normal hierarchy and 0.12 eV in the inverted one; distinguishing these cosmologically requires measuring the total mass to a 5-sigma precision of 0.06 eV. This is typically done by comparing the amplitude of clustering at $z = 1000$ from CMB observations to that in the late-time evolved universe. Because of parameter degeneracies, achieving such tight constraints will also require significant improvements in constraints on the optical depth to reionization, τ. These are obtainable from next-generation large-scale ($\ell < 30$) CMB E-mode polarization measurements and potentially also from small-scale kinematic SZ measurements and measurements of the Dark Ages. Improving the measurement of τ to the required precision of 0.002 (1σ), near the CMB cosmic variance limit, in the next decade will be vital for achieving precise constraints on the neutrino mass hierarchy. Similarly, the low-redshift amplitude of clustering needs to be measured to this level; this can be done with the funded upcoming facilities that will measure gravitational lensing of both galaxies and the CMB, cluster abundances, the intergalactic medium, and redshift-space distortions.

C-Q3c. End-to-End Tests of Cosmology

Diverse and precise cosmological measurements that probe multiple epochs of cosmic history will allow for stringent cross-epoch end-to-end tests of cosmology, enabling explorations of the consistency of cosmological models across cosmic time and providing ways to challenge the standard cosmological model. To maximize the power of these tests, measurements at low redshift will need to reach a precision comparable to those from the early universe, notably from observations of the CMB. As an example, at the precision of today's (2020) surveys, observations have revealed a growing discrepancy between local and CMB-epoch calibrations of the Hubble constant. While this may be a sign of unaddressed biases in the measurements, it could also be an indication of new physics beyond the standard cosmological paradigm. These tests concern comparison of absolute distance scales in the early and late universe and require improved measurements across cosmic time: local expansion rate measurements at low redshift; galaxies, quasar, and Lyman-alpha forest BAO measurements, strong lensing cosmography, supernovae, and gravitational wave standard siren measurements at intermediate redshifts; and small-angle CMB measurements at high redshift. If tensions persist, efforts will be needed to probe scales that can distinguish between changes in the early expansion rate and the speed of sound in the primordial plasma. Such measurements will be accessible to probes of the Dark Ages discussed in the discovery area below.

Another ongoing area of work is to compare the amplitude of structure fluctuations at low and high redshift. Although this was already mentioned in the context of neutrino masses and modified gravity explanations of cosmic acceleration, the evolution of large-scale structure could reveal other extensions in the dark sector, such as late-time decaying particles or time-varying masses.

Summary of Capabilities Needed for C-Q3

To complement the funded upcoming large survey facilities targeting cosmic expansion, next-generation high angular resolution CMB experiments to measure lensing and wide-angle CMB polarization maps to improve the measurement of the optical depth to reionization are needed. Support is further needed for facilities for spectroscopic training of photometric redshifts for weak lensing samples, follow-up of Type Ia supernovae, and narrow-field instruments for local distance measurements and strong lensing cosmography, such as could be provided by large-aperture optical telescopes. The panel anticipates that the early returns from the funded upcoming facilities may be transformational, and a mid-decade assessment will be important to shape plans for future investments in this area, including opportunities to tune the observing plan for Roman Space Telescope.

C-Q4. HOW WILL MEASUREMENTS OF GRAVITATIONAL WAVES RESHAPE OUR COSMOLOGICAL VIEW?

Just 10 years ago, the Astro2010 Cosmology and Fundamental Physics panel listed gravitational wave astronomy as its discovery area. In the intervening decade, LIGO has observed the merger of tens of binary black holes, and its discovery of a binary neutron star merger has heralded the era of multimessenger astronomy. In the next decades, gravitational wave measurements will span a wide range of frequencies, from nHz with pulsar timing arrays, to mHz with LISA, to kHz with ground-based instruments. This new astronomical window will open a large dynamic range of time and scale for cosmological inferences, from a potential stochastic gravitational background originating in the early universe, to new particles created in the vicinity of rotating black holes, to the current expansion rate of the universe. Indeed, our understanding of the potential reach of these observations is still maturing, and they may take us in unexpected directions. While some of the topics described below overlap aspects of

the previous questions, the panel believes that this new and rapidly expanding view of the universe offers an opportunity for cosmological discovery that needs to be specifically recognized.

C-Q4a. The Stochastic Gravitational Wave Background

The detection in the nHz to kHz bands of a stochastic gravitational wave background beyond that expected from compact-object sources would provide a unique window into the thermal history of the universe during otherwise inaccessible times. As the universe cools, phase transitions and their associated topological defects could produce gravitational waves. Gravitational waves far more intense than those expected from simple inflation models could also arise from the start of the Hot Big Bang, and comparison to the ultra-low-frequency waves being sought in CMB large-angle polarization measurements would probe a large lever arm in the spectrum of gravitational waves. Pulsar timing arrays and gravitational wave detectors are sensitive to this stochastic background, provided that one can isolate the background signal from that of nonprimordial compact object mergers.

C-Q4b. Standard Sirens as a New Probe of the Cosmic Distance Scale

With the discovery in 2017 of both gravitational waves and electromagnetic signatures from a binary neutron star merger, scientists were able to make the first “standard siren” measurement of the cosmic distance scale. This measurement was enabled by the exquisite predictive power of GR, which allows the gravitational luminosity distance to be directly measured for individual compact object mergers. This method has notable advantages, such as its reliance on laboratory calibration and independence from the effects of astrophysical dust, although, like standard candle methods, it suffers at high redshift from magnification uncertainties from gravitational lensing. In the coming decades, standard siren samples will increase enormously in size and quality. Using low-redshift mergers, LIGO and Virgo may provide an independent assessment of the current Hubble constant tension to the 1 percent level. Using supermassive black hole binaries, LISA will reach out to redshift ~10, providing a means to build a single distance scale over a remarkable span of cosmic history. Eventually, using the individual events to build maps may enable novel cosmological tests.

Electromagnetic counterparts will be crucial for many applications and will require extensive observing resources to locate and study. The panel expects that the gravitational wave network sensitivity will soon place substantial demands on the capacity of the observatories needed to find the electromagnetic counterparts and acquire the source redshifts. Continued access to follow-up resources will be important to speed progress using this probe.

C-Q4c. Light Fields and Other Novel Phenomena

The increased precision and expanded frequency range of future gravitational wave facilities will offer numerous opportunities to uncover novel phenomena that are now only theoretical speculations. To list some cosmological examples: (1) Gravitational wave emission from rapidly spinning black holes may reveal light bosonic particles whose Compton wavelength matches the horizon size. These bosons need not be a major component of dark matter, but their existence would be an intriguing clue to other light states. (2) Mergers might be found from sub-stellar mass black holes or from extreme redshift, suggesting a new cosmological source of compact objects. (3) Waveforms of merger events might display signatures of gravitational lensing by dark matter substructure. (4) Study of compact object mergers might even reveal a new aspect of strong-field gravitational physics that could perhaps be connected to cosmic acceleration. We should be prepared to be surprised when looking through the gravitational wave window into the unseen relativistic universe.

Summary of Capabilities Needed for C-Q4

Capabilities needed include improvements in gravitational wave detection, particularly through deployment of LISA, coupled with increasing efforts in multiwavelength electromagnetic and multimessenger studies to characterize the population of merger events.

DISCOVERY AREA: THE DARK AGES AS A COSMOLOGICAL PROBE

Our understanding of how the universe began is measured through the fingerprints that the Hot Big Bang left in matter density fluctuations. Unfortunately, these fingerprints are often smudged. At small scales, Silk damping in the CMB and the nonlinear astrophysics of galaxies hides and confuses the primordial density fluctuations, while at large scales measurements are limited by the volume of space that can be observed with galaxy surveys or on the surface of last scattering. At the end of the Dark Ages, the neutral hydrogen pervading the universe became visible against the CMB backlight, enabling observations of the primordial density fluctuations over a vastly larger range of scales than for any other cosmological probe. The panel sees 21 cm and molecular line intensity mapping of the Dark Ages and reionization era as both the discovery area for the next decade and as the likely future technique for measuring the initial conditions of the universe in the decades to follow.

C-DA1. The End of the Dark Ages

As the first luminous objects formed, it is expected that the vast majority of the baryons were very cold and that the hydrogen spin temperature was in equilibrium with the CMB. The first Lyman-alpha photons then coupled the hydrogen spin temperature to the gas temperature, highlighting the neutral hydrogen against the CMB. As illustrated by the theoretical work interpreting the surprisingly large global absorption signal detected by the EDGES experiment, the time evolution and spatial fluctuations in this spin temperature provides a wealth of cosmological information.

During the Dark Ages, astrophysical structure formation responded dramatically to the dark matter power spectrum at the smallest scales. Models of dark matter that suppress small-scale power (e.g., warm dark matter or dark sector interactions) delay the formation of the first luminous objects, while theories that enhance small-scale power (e.g., primordial black holes, dark sector interactions that boost early black hole formation, or models that create compositional fluctuations on small scales) advance this timing. By using the emergence of luminous systems as a timestamp, cosmologists can leverage the astrophysics of early structure formation to probe the primordial power spectrum at currently inaccessible scales.

Unlocking this cosmological window will require advances in both measurements and theory, but it appears attainable in the coming decade. A goal for the coming decade is reconnaissance across a wide range of redshift, primarily with next-generation interferometric mapping supported by global single-receiver measurements, in order to map the temperature history of the intergalactic gas. While small changes in the timing of galaxy formation can be caused by astrophysical details, large changes in when structure formed would be a hallmark of new physics in the dark sector. As our understanding of reionization and the late Dark Ages improves, we will increasingly be able to disentangle the astrophysics of reionization from effects of cosmology.

C-DA2. The Future of Primordial Density Mapping

As described earlier, understanding the initial conditions of our universe requires observing the primordial density fluctuations. Intensity mapping of neutral hydrogen has the potential to measure these

fluctuations with unprecedented precision and reach. Prior to the onset of nonlinear collapse and galaxy formation, the primordial density fluctuations can be measured to much smaller scales than possible at later times. Further, one can measure far more modes with hydrogen intensity mapping than in the CMB primary anisotropies, as one is no longer limited by the Silk damping of the CMB and one is using a three-dimensional map rather than an angular map. A window for this direct mapping of precollapse structure is predicted to exist at redshift 50. Here, increases by factors of over 100 in scale and a billion in number of modes might be available, giving intensity mapping the potential to provide the next major leap in the understanding of the initial conditions imprinted on the primordial density fluctuations by inflation.

But intensity mapping measurements are in their infancy, and the most ambitious program at redshift 50 requires space-based measurements. While significant technical progress has been made toward the first line measurements of the power spectrum at high redshift, the state of the art is still decades away from superseding the CMB in scientific reach. As in the cases of the CMB, gravitational waves, and weak lensing, the development of intensity mapping from concept to a robust cosmological tool will take several decades of steady support.

In the coming decade, the panel anticipates that neutral hydrogen intensity mapping will mature to the point that it can make the first anisotropy measurements of reionization. This is a crucial milestone, and measuring the process of reionization and the CMB optical depth will improve the current understanding of cosmology. The panel also hopes to see the first measurements of the BAO scale using either the 21 cm or other atomic or molecular emission lines. As these techniques mature, the panel expects the precision, angular scale, and redshift of the measurements to steadily improve. A 30- to 40-year goal would be to map the density fluctuations in the pre-reionization universe with an unprecedented number of modes traceable to the primordial density fluctuations, using the power spectrum and non-Gaussianity to measure the statistical initial conditions of the universe.

Summary of Capabilities Needed for the Discovery Area

Needed capabilities include next-generation 21 cm interferometers targeting both the reionization epoch and lower redshifts, along with planning toward very high redshift mapping. Progress will require both higher sensitivity and a better understanding of instrumental systematics and astrophysical couplings.

CROSS-CUTTING CAPABILITIES

This appendix identifies new observational capabilities needed to address the science questions and discovery area. In addition, the panel identified the following cross-cutting capabilities needed to support the overall cosmological research enterprise.

Tremendous opportunities will be offered by the facilities currently nearing completion. These facilities will produce vast data sets that will be useful across a wide range of efforts, especially when these data sets are combined. Fully leveraging the cosmological utility of these observations will require elaborate analyses and extensive collaboration beyond the scope of an individual investigator grant. The panel is concerned that potential scientific output could be unrealized owing to a lack of available human resources to fully and, where appropriate, collaboratively analyze and exploit data sets. The panel urges that attention be given to the support of these larger analysis efforts.

Computational and theoretical studies of cosmology are critical to support the field. The impact of modern computing on cosmology is ubiquitous—ranging from ambitious data reduction methods, to detailed statistical analyses, to high-performance simulations—while theoretical research continues to contribute important new physical hypotheses to be tested as well as calculational and statistical opportunities to extend methods for interpreting the complex data sets derived from both observations and

simulations. Cosmology is particularly remarkable within astrophysics, as it is often concerned with testing theoretical models that provide specific and realistic physical initial conditions, such that forward simulation to compare to observational results is an essential aspect of the interpretation of cosmological data. Advances in theory and scientific computing (the latter enhanced by advances from the data science and machine learning communities) directly enable analyses from current experiments and help to guide the design of future ones. A commitment to the public release of both data and analysis software from next-generation projects, as well as to the development of software that makes good use of the computational power provided by new computer hardware architectures and facilities, will continue to push the state-of-the-art in these areas.

Last, today's cosmological experiments and facilities rely on technology far beyond what was available in past decades, and the health of the field surely depends on continuing this technological growth. Whether for detectors, correlators, robotic mechanisms, novel optics, or technologies enabling cheaper and more capable spaceflight, pushing the state of the art in cosmology requires strategic support for technology development projects.

CONCLUSION

The origin, composition, and physical laws of the universe are ancient sources of wonder and ever-present drivers in our study of astronomy. The coming decade will be a bold new chapter in that cosmological story, with ambitious facilities and unprecedented data sets uniting with powerful statistical, computational, and analytical methods to explore many different frontiers.

This panel has identified four critical science questions and one discovery area that it believes are ripe for substantial progress this decade: (1) What set the Hot Big Bang in motion? (2) What are the properties of dark matter and the dark sector? (3) What physics drives the cosmic expansion and large-scale evolution of the universe? (4) How will measurements of gravitational waves reshape our cosmological view? The discovery area is the Dark Ages as a cosmological probe. These and their parts are summarized in Box C.1. Table C.1 presents the highest profile yet-unfunded capabilities needed to address the cosmology science questions and discovery area. These questions build on the successful framework of the standard cosmological model to search for distinctive signatures from the dark sector, the early universe, the cosmic expansion history, the gravitational wave window, and the Dark Ages, all of which can reveal rich new phenomena in realistic physical theories. Through the exquisite experiments, observations, and computations now possible, we can explore domains in energy, space, and time previously inaccessible yet critical to understanding our place in the universe.

BOX C.1
Summary of Science Questions

C-Q1: What set the Hot Big Bang in motion?	C-Q1a: Primordial gravitational waves C-Q1b: Non-Gaussianity of the large-scale structure of the universe C-Q1c: The initial power spectrum of density fluctuations
C-Q2: What are the properties of dark matter and the dark sector?	C-Q2a: Dark sector signatures in small-scale structure C-Q2b: Dark sector imprints on Big Bang nucleosynthesis and recombination C-Q2c: Annihilation by-products
C-Q3: What physics drives the cosmic expansion and large-scale evolution of the universe?	C-Q3a: The physics of cosmic acceleration C-Q3b: The properties of neutrinos C-Q3c: End-to-end tests of cosmology
C-Q4: How will measurements of gravitational waves reshape our cosmological view?	C-Q4a: The stochastic gravitational wave background C-Q4b: Standard sirens as a new probe of the cosmic distance scale C-Q4c: Light fields and other novel phenomena
Discovery Area: The Dark Ages as a cosmological probe	C-DA1: The end of the Dark Ages C-DA2: The future of primordial density mapping

TABLE C.1 Capabilities Needed to Address the Cosmology Science Questions and Discovery Area

Capability	Science Enabled	Future Needs
Wide-angle CMB polarization mapping	(C-1a) Primordial gravitational waves; (C-3b) neutrino mass from E-mode optical depth measurement; (C-3c) end-to-end test of LSS growth	Reach detection threshold of r ~ 0.001; measure optical depth to recombination to 0.002 (1\sigma).
Arcminute-scale CMB mapping	(C-1a) Primordial gravitational waves delensing; (C-1b) non-Gaussian LSS using kinematic SZ field; (C-1c) deviations from power-law adiabatic fluctuations; (C-2b) measurement of relic radiation density; (C-3a) CMB primary anisotropies and lensing to study dark energy; (C-3a) thermal SZ and CMB lensing for cluster cosmology; (C-3b) kinematic SZ study of reionization epoch; (C-3c) end-to-end tests of large-scale cosmological model	Approach cosmic variance limit of primary ($\ell < 4000$) anisotropies; most-sky delensing maps for r ~ 0.001; $\sigma(N_{eff})$ ~ 1 percent of neutrino density.
Spectroscopic large-scale structure	(C-1b) Non-Gaussianity; (C-1c) deviations from power-law adiabatic fluctuations; (C-3a) acoustic scale measurements; (C-3a) dense redshift and lensing survey for LSS growth history; (C-3b) neutrino mass from low-redshift LSS amplitude; (C-3c) end-to-end tests of large-scale cosmological model	$\sigma(f_{NL})$ ~ 0.2; amplitude of structure σ ~ 0.2 percent.
Pulsar timing	(C-2a) Dark sector small-scale structure; (C-4a) stochastic gravitational waves background	Next-generation radio telescopes for pulsar timing.
Narrow- and moderate-field, high-sensitivity, high-multiplex spectroscopy	(C-2a) Dark sector small-scale structure from strong lenses and dwarf galaxy mass profiles; (C-3a) expansion history from strong lensing time delays; (C-3b) spectroscopic photometric redshift training for measurements of structure growth from weak lensing; (C-3c) end-to-end tests of large-scale cosmological model	Next-generation large-aperture OIR telescopes with integral-field or high-multiplex spectrographs.
21 cm interferometers	(C-2a) Dark sector small-scale structure; (C-3b) support modeling of CMB optical depth; (C-DA1) unusual IGM temperature histories; (C-DA2) primordial density mapping	Long-term, map at z > 50; decadal-scale, map at reionization epoch and lower redshifts.
UV/Optical high-dispersion spectroscopy	(C-2b) BBN light element abundances	Next-generation large-aperture OIR telescopes with high-dispersion spectrographs.
TeV imaging	(C-2c) Search for TeV WIMP annihilation	Reach 1×10^{-28} cm^3/s cross section at 2 TeV.
GeV imaging	(C-2c) Source of GeV excess in Milky Way center	Improve angular resolution and/or sensitivity.

Capability	Science Enabled	Future Needs
Time-domain follow-up	(C-3a and C-3c) Cosmic expansion history and H_0; (C-4b) standard sirens; (C-4c) novel cosmological gravitational wave phenomena	Spectroscopy and imaging for supernovae follow-up, gravitational waves counterparts, and strong lensing cosmography; improved pan-chromatic sensitivity and access.
Local distance measurement	(C-3a and C-3c) Hubble constant	For example, next-generation large-aperture OIR telescopes.
Gravitational wave detection	(C-3a and C-4b) Standard sirens for cosmic expansion history and H_0; (C-4a) stochastic gravitational wave background; (C-4c) novel cosmological gravitational wave phenomena	Next-generation pulsar timing; terrestrial detectors not in Astro2020 scope.
Large-scale computation; theory research; technology development; large data set analysis; sharing and curation of software and data sets	Ubiquitous contributions	These cross-cutting capabilities will require consistent attention and funding.

NOTE: This table of capabilities focuses on the highest profile yet-unfunded items, echoed from the summaries at the end of each section. Other capabilities are listed in the text. Needs provided by facilities existing or currently under construction are not included. Some capabilities are described in terms of the observational goal, agnostic to wavelength, as there are multiple plausible paths. The panel stresses that neither these capabilities nor the science questions are presented in priority order.

D

Report of the Panel on Galaxies

FRAMEWORK

Galaxies are the basic unit of observable structure on cosmic scales, themselves residing within a hierarchy of groups, clusters, and superclusters, and displaying an astonishing diversity of properties. The luminous regions of galaxies are vast, but the dark spaces between them are much vaster still. We now understand that this apparent emptiness is partly an illusion. The dark matter halos of galaxies extend to great distances and meld into a filamentary cosmic web. The dark matter structures are permeated by a tenuous circumgalactic, intracluster, and intergalactic medium, made of primordial hydrogen and helium that may eventually join galaxies and form into stars, as well as of enriched gas that carries the products of previous stellar generations back into intergalactic space.

The goal of the field of galaxy formation and evolution is to achieve a predictive formulation of the assembly histories of galaxies and their dark matter halos, together with the evolution of their stellar populations, black holes, physical structures, chemical content, and circumgalactic, intracluster and intergalactic media. Observations characterize both the common trends and the diversity of galaxy properties, and their evolution with time. Theoretical models aim to explain these observations with a priori physics while providing predictions that can be tested with new data.

Over the past two decades, enormous progress has been made in linking galaxies and larger baryonic structures to their dark matter halos and in understanding the processes responsible for that link. We know that the luminous bodies of galaxies are part of an interconnected ecosystem that includes their surrounding medium out to intergalactic scales. The flow of matter and energy throughout the entire ecosystem is likely responsible for both the diversity and regularity of galaxies. Stars and black holes, the prime engines of the matter and energy flow, are believed to have powered the major phase transition—cosmic reionization—that the universe underwent in the relatively short period between a half and one billion years after the Big Bang.

These newly established paradigms point to critical, and addressable, paths forward. We must observe and understand the sources that caused cosmic reionization, and we must isolate the individual physical processes that drive the evolution of the ecosystem and govern the connection between gas, stars, black holes, galaxies and their dark matter halos. These challenges can be addressed only with coupled advancements in multiwavelength and multiscale observations, models, and simulations, while leaving room for the serendipity that has often led to major discoveries and leaps in understanding. The four science questions formulated in this report, while not exhaustive of this rich field, are the most compelling to address in the 2020s and beyond. These questions are expected to experience major advances over the next decade thanks to upcoming powerful space and ground facilities, and to the increasing sophistication of simulations and models; but they also require new, future capabilities. The discovery area highlights a new observational approach that is becoming technologically feasible for the first time and will produce a more complete picture of the baryons in the universe.

> Question D-Q1. How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?

Question D-Q2. How do gas, metals, and dust flow into, through, and out of galaxies?
Question D-Q3. How do supermassive black holes form and how is their growth coupled to the evolution of their host galaxies?
Question D-Q4. How do the histories of galaxies and their dark matter halos shape their observable properties?

Discovery Area: Mapping the circumgalactic medium and the intergalactic medium in emission

STATE OF THE FIELD

Galaxies are open systems with extensive circulation of energy, gas, and metals between their stellar bodies and the surrounding circumgalactic medium (CGM), and, farther out, the intracluster and intergalactic media (ICM and IGM). The mechanisms that drive this circulation take the catch-all name of "feedback," which describes the energy, momentum, and matter ejection driven by star formation (e.g., stellar winds, radiation pressure, supernova explosions) and by accreting supermassive black holes (SMBHs). The competition between accretion and these feedback processes regulates the growth of galaxies, ultimately accounting for both their large diversity and common trends. Much effort over the past decade has been devoted to investigating the "what, when, and where" of the regulatory mechanisms of galaxies. A relatively simple model in which a central galaxy's gas accretion tracks its halo's dark matter accretion, and the efficiency of conversion to stars depends on the potential well depth, can account for many observed galaxy properties over a wide range of redshift.

The First Structures and Reionization

The first baryonic structures—stars, black holes (BHs), and galaxies—arise within the first 0.5 Gyr of cosmic history ($z > 9$). By $z \sim 6$ (age ~ 1 Gyr), the UV photons from these systems have reionized intergalactic hydrogen throughout the universe. Theoretical models predict a patchy reionization process in which ionized bubbles gradually expand and overlap, and quasar absorption spectra provide tentative evidence for this inhomogeneous structure at $z \sim 6$–7. For galaxies to be the principal sources of reionization, significant fractions of ionizing photons would need to escape low-mass ($M_{stars} < 10^8\,M_\odot$), and currently unobservable, galaxies. Rare SMBHs with masses $\gtrsim 10^9\,M_\odot$ are already present by $z \sim 6$. Growing BHs to such high mass within a Gyr of the Big Bang requires either massive initial seeds ($M \sim 10^4$–$10^6\,M_\odot$) produced by exotic physical conditions or highly efficient accretion from stellar mass seeds—or both. Understanding the origin of the early SMBHs and the contribution of radiation from accreting BHs to reionization remain open questions both observationally and theoretically.

The Growth of Galaxies and Black Holes

After the conclusion of reionization, structures appear to grow in a mostly self-regulated equilibrium mode, where the stellar mass, the molecular gas reservoir, and the star formation rate (SFR) of galaxies track each other across cosmic time. The fraction of mass in galaxies that is in cold neutral interstellar gas decreases steadily with time from >80 percent at $z > 3$–4 to ~10 percent today, supporting the direct link between baryonic accretion onto galaxies and stellar mass build-up. However, the formation and evolution of the Hubble sequence of disks, bulges, and spheroids is still poorly understood. Early disk formation appears to be chaotic, followed by "disk settling," where disk galaxies evolve from morphologically disturbed and clumpy systems with predominantly disordered gas motions to thin ordered disks over the past ~10 Gyr. Theory and observations suggest that secular disk instabilities, gas accretion, feedback, and galactic mergers all contribute to bulge formation, but their relative importance

and contribution to the diversity of today's bulges are still to be established. One striking puzzle is the emergence by $z \sim 4$ of massive evolved spheroids with little ongoing star formation. While the stellar mass of central galaxies is tightly correlated with the halo mass, star formation in satellite galaxies appears to become quenched after the stars enter the parent galaxy's halo, which suppresses their stellar mass growth and reddens their colors relative to isolated systems. The observed correlation between SMBH mass and host galaxy stellar mass, well established for galaxies with $M_{stars} > 10^{11} M_{\odot}$, hints at coupled evolution, with BH growth regulating stellar mass or vice versa. The integrated rate of SMBH accretion over cosmic time tracks the cosmic SFR with a volume-averaged ratio that has remained broadly constant over the past 10 Gyr. However, beyond $z \sim 2$, much of our knowledge of SMBHs is confined to the luminous quasars, leaving a gap in our understanding of galaxy-SMBH coevolution at lower luminosity.

The Ecosystem: Flows in the Circumgalactic, Intracluster, and Intergalactic Media

Decoding the multiple processes that connect galaxy growth, star formation, SMBH accretion, and the CGM has been a central concern over the past decade, with advances on many fronts. Observations reveal ubiquitous outflows from typical star-forming galaxies at $z \sim 2$–4 and from rapidly star-forming galaxies at low redshift. It remains unclear, however, how much of the ejected material cycles back into galaxies and how much remains in the halo or escapes to the ICM/IGM. CGM observations—principally X-ray emission and UV absorption—show that hot ($T \sim 10^6$–10^7 K) and cool ($T \sim 10^4$–10^5 K) gas coexists in the halos of many galaxies and that the metal content of the CGM is highly inhomogeneous, with some metal enriched pockets that could exceed the metallicity of the central galaxy's stars and ISM. Galactic winds contain co-existing gas phases from $T \sim 10^7$ K plasma down to $T \sim 10$ K molecular gas. On larger scales, the deepest integral field unit (IFU) observations and "tomographic" absorption maps toward dense grids of background sources are providing the first hints of the filamentary IGM structures predicted so vividly by simulations.

The Role of Feedback in Powering and Feeding the Ecosystem

Explaining the stellar masses and the mass-metallicity-SFR relation of galaxies requires characteristic mass-loading factors (ratio of outflow rate to SFR) of order unity for Milky Way-like galaxies with halo mass $M_{halo} \sim 10^{12} M_{\odot}$, rising to much higher values of 10–50 for dwarf galaxies in low-mass halos. Stellar feedback, which has been shown to correlate with the surface density of star formation, is considered the dominant mechanism for driving outflows in low- and intermediate-mass galaxies, and, by driving ISM turbulence, it regulates the efficiency of star formation on scales from individual molecular clouds to entire galaxies. For massive galaxies and halos, there is broad consensus that radiation and jets powered by SMBH accretion play central roles in suppressing cooling from the CGM and quenching star formation, although much is still unclear on how these processes work. Observations of the past decade provide strong evidence for the effects of SMBH feedback on intracluster gas, connecting phenomena more than nine orders of magnitude in spatial scale.

Stellar and SMBH feedback, however, may not be sufficient to account for all observed galaxy properties. Questions exist on the galaxy-CGM-feedback coupling mechanisms, on the origin of transition points (e.g., the maximum baryon conversion efficiency at $M_{halo} \sim 10^{12} M_{\odot}$) and on the normalization, scatter, and evolution of galaxy scaling relations. For $M_{halo} > 10^{12} M_{\odot}$, environmental quenching mechanisms that prevent the gas from accreting and/or cooling onto a galaxy and forming stars efficiently may need to be present in order to explain the presence of massive quiescent spheroids at high redshift. In low-mass galaxies, $M_{halo} < 10^{10} M_{\odot}$, simulations show that stellar feedback needs to be coupled with the cosmic UV background to inhibit star formation and explain their low gas-to-star conversion efficiency.

Zooming into the Physics: The Milky Way and Nearby Galaxies

Nearby galaxies and the Milky Way offer testing grounds for understanding structure formation and evolution that are unparalleled in resolution and detail, and for isolating the dominant physical mechanisms that drive them. Resolving the stellar populations of nearby galaxies and their satellites has enabled reconstructing much of their accretion histories. The ultra-faint dwarf satellites of the Milky Way and M31 probe the low-mass threshold of galaxy formation, with stellar populations that may have formed before reionization. Dynamical measurements of these galaxies show many orders of magnitude range in stellar mass over a narrow range of halo mass, $M_{halo} \sim 10^8$–10^{10} $M_\odot$, suggesting that the final stellar masses are sensitive to intersections between star formation histories, feedback processes, dark matter assembly, and, possibly, dark matter physics, in ways that are still not fully understood. Large populations of "ultra-diffuse galaxies," with low central surface brightness and large sizes, have been discovered and characterized in galaxy groups and clusters. They exhibit remarkable properties, from exceedingly low-velocity dispersions—suggesting little dark matter within their optical radii—to anomalous globular cluster (GC) systems.

Within the Milky Way, Gaia measurements of stellar distances and proper motions, multi-element spectroscopic surveys of hundreds of thousands of stars across all components of the Galaxy, and asteroseismic measurements that have opened entirely new routes to determining stellar ages are driving progress with samples of millions of stars with accurate radial velocities and metal abundances. These chemodynamical measurements provide increasingly strong evidence that many stars in the disk of the Milky Way migrate far from the radius at which they were born. They have also revealed that much of the Milky Way's inner stellar halo was contributed by a single dwarf galaxy merger in the distant past, and that disk star kinematics are perturbed by the Sagittarius dwarf and other satellites.

The Backbone: Successes and Challenges for Theory and Simulations

Simulations of galaxy formation and evolution have made enormous progress over the past decade, and now routinely produce galaxies with global and structural properties that match many key observations. While these simulations begin with well-defined cosmological initial conditions, small-scale phenomena, including star formation, accreting SMBHs, and their feedback, are not resolved and are therefore implemented via sub-grid models. Despite significant differences in the sub-grid prescriptions, different simulation suites predict similar results for statistical properties, global scaling relations, and, qualitatively, some morphological features. But it remains unclear how the physical processes that occur on the parsec (pc) and sub-pc scales of gas clouds, individual stars, supernovae (SNe), and SMBHs couple across decades in time and space, leading to the observed properties of galaxies and their surrounding media. There is also no consensus on the minimal or essential set of physical processes that must be included in galaxy evolution models. Furthermore, because the sub-grid models are tuned to match specific statistical galaxy properties (such as the stellar mass function and the mass-metallicity relation), their predictive power is currently limited. A key challenge for the next generation of simulations will be confronting them with observations that they were not tuned to reproduce, such as CGM observations and detailed galaxy morphologies and gas contents over a wide range of redshift.

D-Q1. HOW DID THE INTERGALACTIC MEDIUM AND THE FIRST SOURCES OF RADIATION EVOLVE FROM COSMIC DAWN THROUGH THE EPOCH OF REIONIZATION?

As the first sources formed, the universe emerged from the dark ages and became progressively transparent to photons. How this happened is a central question in modern astrophysics research. It is

expected that a significant number of long-standing puzzles about the first billion years will be answered in the 2020s. However, the ultimate frontier of the ignition of the first stars and seeds of the first SMBHs will likely remain *terra incognita*. Beyond the next decade, the goal of the 2030s will be to discover, determine, and interpret the properties of the very first stars, galaxies, and BHs, together with detailed studies of typical Milky Way-progenitor galaxies at $z > 10$. While the topic of reionization is covered in Appendix C as well, this appendix concentrates on the properties of the sources of reionization and the structures that emerge from this epoch.

D-Q1a. Detailed Thermal History of the Intergalactic Medium and the Topology of Reionization

Intergalactic gas remains cold and opaque during the dark ages. As soon as the first sources form, light from these sources, in the form of high-energy photons, is expected to convert the cold gas to ionized plasma. The photons inject heat into the IGM and allow light to travel freely through intergalactic space. This reionization process is observed to unfold rapidly and is completed in the first billion years. One of the puzzles is how reionization occurs, including the identification of the dominant sources of ionizing photons during the epoch of reionization (EoR; from first light to $z \sim 6$). Because the thermal and ionization histories of the universe are intimately coupled, a deeper understanding can be gained through measurements of the timeline, thermal history, and topology of reionization. These measurements require improved HI line intensity maps (which measure the redshifted HI 21 cm transition) and deep wide-field optical/near-IR imaging surveys covering the EoR. The temperature evolution from HI alone will help discriminate among the sources of heating, possibly distinguishing stellar-origin and heavy BH seeds. The topology, or distribution of angular sizes and clustering of the ionized bubbles, will test the nature of the sources of ionizing photons. Cross-correlation of HI maps with high-redshift sources (and their properties) will be a powerful diagnostic of how the heating/ionization of the IGM progressed, and also of the nature of the dominant sources. The HI observations need to span tens of square degrees with noise levels of 0.2 mK^2 at $z = 8$ and 100 mK^2 at $z = 15$. Wide-field imaging surveys (e.g., with the Roman Space Telescope) will observe $>10^5$ galaxies out to $z \sim 10$–12, down to $K_{AB} \sim 26$, suitable for cross-correlation with the HI signal. Accurate galaxy redshifts will be key for taking full advantage of the correlation signal and will require highly multiplexing near-IR spectrographs on telescopes that are about two orders of magnitude more sensitive than the 10 m class telescopes currently available. This increase in sensitivity can be achieved with a combination of larger size and higher angular resolution.

Targeting highly ionized patches during the reionization epoch provides a unique window into the process of reionization and the nature of ionizing sources. In the vicinity of quasars, intense radiation powered by the central SMBHs forms isolated ionized bubbles embedded in an otherwise mostly neutral IGM. High-resolution absorption spectroscopy of the most distant quasars enables detailed density and temperature maps both inside and outside the ionizing bubbles around the quasars. Roman will detect ~2600 $z > 7$ quasars, and ~20 percent of them will be at $z > 8$; Euclid is expected to discover ~150 bright ($J_{AB} \lesssim 22$) QSOs at $z \sim 7$–9. Medium spectral resolution ($\delta v < 50$ km/s) near-IR spectroscopy is required to resolve the Lyα line in the ionizing bubbles; high-resolution spectroscopy ($\delta v < 10$ km/s) will resolve transmission spikes that arise in low-density, more transparent patches in the neutral IGM. These require echelle spectrographs to observe objects that are two orders of magnitude fainter than achievable with current telescopes.

D-Q1b. Production of Ionizing Photons and Their Escape into the Intergalactic Medium

We do not know what the dominant sources of reionization are and how their relative importance changes at different stages of the EoR. Current evidence points toward low-mass galaxies at late stages ($z \sim 6$–9**)**, although active galactic nuclei (AGNs) may play an important role as well. Establishing the role of low-mass galaxies in the EoR requires determining (1) the number of low-mass galaxies in the early

universe, and (2) the amount of ionizing photons they produce that escapes into the IGM (i.e., Lyman continuum, or LyC, escape fractions). The first goal requires measurement of the faint ends of UV galaxy luminosity functions, which may extend into the regime of early GCs or proto-GCs, during the EoR. Pioneering studies have spectroscopically confirmed the redshifts of a few bright galaxies ($H_{AB} \sim 25$–26) at $z \sim 7$–9, but hundreds of fainter galaxies at these redshifts must be confirmed. Ultradeep imaging surveys (with, e.g., the James Webb Space Telescope [JWST]) are needed, assisted by magnification from foreground galaxy clusters. A faint galaxy at $z = 8$–20 of $M_{UV} = -12$ corresponds to $K_{AB} = 35$–36.5. Detection and confirmation of such a galaxy requires both near-IR capabilities two orders of magnitude more sensitive than currently available and magnifications ~10–100.

Direct measurements of the LyC escape fractions of galaxies in the reionization era are not possible, owing to absorption by the intervening IGM, but are possible at lower redshifts. Therefore, the production and escape of ionizing photons from sources between first light and $z \sim 6$ need to be inferred by comparing the internal properties of the sources to those of galaxies at much lower redshift for which LyC escape fractions can be measured (see Question D-Q4c). This requires measurements of gas kinematics, gas conditions, geometries, and chemical compositions of the sources, as well as maps of gas inflows and outflows. JWST will reach depths of $K_{AB} \sim 30$–32, and therefore will detect bright galaxies to $z \sim 14$–16, albeit in small arcminute-size fields. However, most galaxies at $z > 10$ are low-mass and extremely compact (<<1 kpc), and their lines are expected to be narrow ($\delta v \ll 100$ km/s). Therefore, near-IR integral field unit (IFU) spectroscopic capabilities that deliver spatially resolved (~100–200 pc) information at medium spectral resolution (R ~ 3000–5000) for $K_{AB} \sim 32$ are needed for mapping the low-mass galaxies. Even in the presence of lensing, this requires facilities that have at least two orders of magnitude higher sensitivity than the ones currently available combined with an angular resolution capable of resolving ~100 pc at $z = 14$–16. Measurements of several hundreds of galaxies across the luminosity function, down to $M_{stars} \sim 10^5$–$10^6\ M_{\odot}$, are needed to understand LyC photon production and escape. This requires large fields-of-view or multiple pointings, as at $K_{AB} = 30$ the density of galaxies is ~30–3000 per deg^2 between $z = 12$ and $z = 8$. The interpretation of this wealth of data will require detailed models and simulations of the properties of galaxies during the EoR in order to understand the conditions for the escape of LyC photons.

D-Q1c. Properties of the First Stars, Galaxies, and Black Holes

Population III (Pop III) stars are the first stars to form after the Big Bang, perhaps as early as $z \sim$ 50–60 in the LambdaCDM model, and are expected to form until $z \lesssim 6$ in isolated regions. Current expectations for Pop III stars are informed by models that include large unknowns such as their initial mass function (IMF), formation mechanisms, evolution, and the environments in which they form and that they impact. This is an area where synergy with observations is expected to spur theoretical developments of models for the formation and evolution of the first stars, BH seeds, and galaxies.

Pop III stars are likely extremely faint and rare (possibly AB~35, but more typically AB~39; sky density~1 Mpc^{-3}), and their direct detection will require more sensitive near-IR capabilities than offered by JWST. The possible exception may be rare but bright rapidly accreting supermassive, cool, red supergiant Pop III stars. Lensing may enable detection of Pop III stars that are extremely massive and/or form in dense clusters. Very rare caustic-crossing events may reach fainter Pop III stars. However, detection requires extensive monitoring of numerous lensing clusters to AB $\lesssim$ 29 mag. Long-lived low-mass Pop III stars ([Z/H] ~ –5), or their direct Pop II descendants ([Z/H] $\gtrsim$ –4), may be observable in the Local Group (V $\lesssim$ 18). Potential targets for such near-field cosmological studies include extremely metal-poor stars in the Milky Way's halo and bulge, old metal-poor GCs, and ultra-faint dwarf galaxies (~10^3–$10^4\ M_{\odot}$). These observations require high-resolution UV spectra (R ~ 30,000) to detect weak metal lines, and dedicated surveys on more sensitive telescopes than currently available. Pushing such studies to statistically significant galaxy samples (~10 Mpc distances) requires even larger facilities (see Question D-Q4b).

Pop III may be observed indirectly via pair-instability supernovae (PISN). Recent models of PISN find that some reach AB ~ 31.5 at z ~ 20, detectable in JWST imaging. However, PISN are expected to be rare (10^{-3}–10^{-2} yr^{-1} $arcmin^{-2}$), and JWST may find only ~5–10 at z > 15 unless a multiyear-wide survey is implemented. Wide-field surveys (e.g., with Roman) can potentially find up to 1000 PISN candidates, but at lower redshifts. Last, models predict that PISN leave a distinctive pattern of elemental abundances, potentially observable in the earliest galaxies and in the ultra-metal-poor stars mentioned above. Abundance measurements may even shine light on the IMF of Pop III stars.

Gamma-ray bursts (GRBs) probe both early star formation, via their measured rates, and the metal enrichment of the IGM, by acting as background light sources for absorption line spectroscopy. GRBs require an alert system and prompt follow-up to identify the ones most likely to be at high redshift, to allow imaging and spectroscopy of the most interesting cases. JWST is limited in its ability to pursue rapid follow-up, although it is a powerful option to determine host redshifts. The path forward requires more agile and high-sensitivity facilities with near-IR imaging and spectroscopy.

D-Q2. HOW DO GAS, METALS, AND DUST FLOW INTO, THROUGH, AND OUT OF GALAXIES?

Pristine gas flows from the intergalactic environment surrounding galaxies, through the galactic halo, to fuel the growth of galaxies. The enriched gas is then returned by galactic winds back into the surrounding diffuse gaseous environment, which can then be accreted back onto the galaxies. The hot (10^7–10^8 K) gas in low-redshift groups and clusters and high-redshift massive clusters has been mapped in emission by X-ray telescopes, with the high angular resolution of Chandra revealing the intricate structures of cool fronts, internal shocks, and buoyant bubbles. However, our empirical constraints on the cooler phases ($<10^6$ K) of the diffuse gas between and beyond galaxies come from absorption-line observations, which probe sparsely distributed individual sightlines, or from stacking and ensemble averages of observations. Given the complexity and multiphase nature of these media, progress in understanding the physical processes that shape the evolution of galaxies and their larger ecosystems hinges on securing multiwavelength maps of gas kinematics, chemical compositions, density, ionization, and thermal structures of individual galaxies, galaxy clusters and their host halos across cosmic time.

D-Q2a. The Acquisition of the Gas Necessary to Fuel Star Formation

As galaxies grow, their reservoirs of fuel for star formation are expected to be replenished. However, little direct observational evidence is available for baryonic accretion at any mass scale or redshift, in part because the accreting gas is expected to be kinematically quiet and often confused with outflow signatures. Further progress in establishing the gas accretion history of galaxies requires spatially resolved imaging spectroscopy of both the diffuse ionized gas and the neutral atomic/molecular gas in galaxy halos. Because of its low column density, diffuse gas has been primarily traced through absorption in the UV and X ray, although more recently emission line measurements of dense, metal-enriched gas as it cools have been possible in the UV/optical. Typical gaseous streams span several hundreds of kpc with velocity dispersion $\delta v \ll 100$ km/s, which sensitive imaging spectrographs would need to match down to sensitivities ~10^{-20} erg s^{-1} cm^{-2} $arcsec^{-2}$ on spatial scales <1 kpc, to track accretion from halo to disk based on the gas kinematics. This sensitivity can be reached only over a small area with current 10 m class telescopes by allowing dedicated week-long integrations; thus, larger facilities are required to map halo-size areas. New sub-grid models calibrated from small-scale observations will need to be developed to provide genuine simulation predictions for all large-scale observable properties of diffuse gas. Such simulations will have sufficient overlap in spatial and temporal scales and in modeled physics with simulations of star formation and ISM to establish self-consistent physical models for different astrophysical phenomena over vast dynamical scales.

Within the disk, the gas is transformed into stars. Observations of a small number of high-redshift massive galaxies have revealed that the cosmic molecular gas content appears to broadly trace the SFR history, while the neutral atomic gas, a precursor of molecules, appears to have evolved more slowly. High-resolution imaging with the Atacama Large Millimiter/Submillimeter Array (ALMA) of far-IR fine structure lines like [CII] is beginning to quantify the ISM in star-forming galaxies, but detailed studies of the cold molecular gas in typical, Milky Way-like or lower mass, galaxies beyond $z \sim 2$ require radio-to-mm observations at sub-kpc resolution and $\delta v \sim 10$–30 km/s in order to resolve and characterize the properties and kinematics of molecular gas clumps with masses ~ a few $\times 10^8 M_{\odot}$. The sensitivities at the required spectral and spatial resolution are at least an order of magnitude beyond those of present-day instruments like the JVLA and ALMA.

D-Q2b. The Production, Distribution, and Cycling of Metals

Heavy elements are synthesized in stars and found in the low-density CGM/IGM far from star-forming regions since early cosmic epochs. However, a complete census of heavy elements in different environments and a robust understanding of the associated distribution mechanisms are both lacking. The presence of heavy elements is expected to alter both thermal and chemical states of the gas. Not only does the cooling efficiency depend sensitively on the gas metallicity, but the formation of molecules and dust grains also correlates strongly with gas metallicity. Tracking heavy elements, from their production in stars, release to the ISM, and escape into the CGM/ICM/IGM provides a complete accounting of these elements, enables identification of dominant enrichment sources, determines the extent of feedback, and constrains the thermal properties of the gas.

Metallicity provides a quantitative measure of the enrichment level of different gas reservoirs. Metallicity measurements of HII regions require understanding of the excitation and ionization mechanisms and observations of weak, narrow ($\delta v < 100$ km/s) diagnostic lines that appear primarily in the rest-frame UV and optical. Sensitive rest-frame UV absorption spectroscopy with $\delta v < 10$ km/s provides a powerful probe of the metal content and homogeneity in the warm (10^4–10^5 K) diffuse CGM/ICM/IGM through observations of ionic absorption features across all redshifts; reaching $UV_{AB} \sim 23$ for background QSOs or bright galaxies would expand from single to multiple lines of sight through an individual halo. High-resolution ($E/\delta E > 1000$) spectroscopy of X-ray emitting gas would provide the necessary constraints for the elemental abundances and abundance patterns in the hot CGM/ICM. Last, wide-field IFUs in the UV and X ray would enable direct mapping of diffuse metal line emission, as described in the Discovery Area section below, and constrain the patchiness of metal mixing in warm phase and hot plasma. Together, these observations, which are beyond what is achievable with current and planned near-term facilities, would enable us to understand how heavy elements are dispersed, ejected, mixed, and redistributed during the lifetime of a galaxy.

D-Q2c. The Coupling of Small-Scale Energetic Feedback Processes to the Larger Gaseous Reservoir

A central puzzle of galaxy evolution is why the measured star formation efficiencies are low, around 10 percent or less, across the entire history of galaxies, even at the peak of their growth ($z \sim 1.5$–4.0) when their gas mass fractions are high, with $M_{gas}/M_{stars} \sim 3$–4. Furthermore, it is unclear why the efficiency depends on the halo mass, with the peak of conversion of baryons to stars in the central galaxy occurring at $M_{halo} \sim 10^{12} M_{\odot}$. Cosmological simulations require feedback from stellar winds/supernovae and from SMBHs at the low and high galaxy mass ends, respectively, to regulate the baryonic accretion onto dark matter halos and reduce star formation efficiency. Models predict that, depending on the halo mass, a catastrophic loss of cool ISM from the galaxy's disk can occur, after which the galaxy remains

quiescent until a new gas supply is accreted. At present, however, these effects of feedback have not been directly observed in situ.

While evidence for star formation driven feedback is widespread, the principal physical mechanisms that drive the multiphase galactic winds are not understood. Establishing how small-scale energetic feedback processes are coupled to the larger gaseous reservoir holds the key for distinguishing between different energy and momentum injection processes, including thermal and nonthermal components such as magnetic fields and cosmic rays, in galaxies of different mass and redshift. It also clarifies when and how the large-scale environment can become hostile to accretion and/or cooling of gas onto a galaxy. Spatially resolved observations of ionized gas within galaxies complement those of the cold neutral and molecular ISM (see Question D-Q2a) and enable a deeper understanding of the formation and survival of dust and molecules. High-resolution, sensitive UV/optical/near-IR spectral maps of galaxies, resolving HII region scales out to $z = 10$ down to line sensitivities of 10^{-19}–10^{-20} erg s^{-1} cm^{-2}, coupled with spatially resolved sub-kpc X-ray, UV, and optical spectral maps of the CGM will establish gas kinematics and chemical imprints from the ISM to the CGM/ICM/IGM and connect small-scale feedback to gas properties on scales out to and beyond the virial radius (~50–300 kpc, depending on redshift). This is beyond what is achievable with current and planned near-term facilities. The interpretation of these observations will require developments in theory and simulations aimed at understanding the physics of feedback and the role of magnetic fields and cosmic rays across multiple scales.

D-Q2d. The Physical Conditions of the Circumgalactic Medium

The CGM lies at the interface between infall and outflows, making it uniquely sensitive to the physics of baryonic flows. In particular, feedback affects the physical state of all baryons within a galaxy's sphere of influence and ensures that a large fraction of both gas and metals associated with the dark matter halos remain in the CGM and local IGM. Theory suggests that, independent of redshift, hot atmospheres develop only in halos $\gtrsim 10^{12}\ M_\odot$, the expected mass threshold beyond which the supply of cool gas to the central regions of the halo is curtailed, slowing the rate of star formation. At higher redshifts, the physical conditions in the CGM record the concurrent global effects of the resultant feedback processes and predict future accretion activity. At low redshifts, the physical conditions in the CGM provide a record of the galaxy's past history of feedback.

The multiphase nature of the diffuse CGM/ICM and its complex dynamics, where all mechanisms are superposed, necessitate a multiwavelength and multiscale approach that includes imaging and spectroscopic studies of individual galaxy halos and clusters over the full spectral range from X ray to radio. The important scales to probe span from large-scale halo environments of ~100 kpc down to star-forming clouds of ~100 pc, while the dynamic range in the gas density and temperature is significantly larger. Current instrumentation does not provide the necessary combination of high spatial resolution, wide field, wavelength coverage, and surface brightness sensitivity to enable such studies. In general, spectroscopic capabilities for resolving narrow ($\delta v < 100$ km/s) kinematic features across the full spectral range are necessary. Wide-field, high spatial resolution IFUs with sensitivities more than 10 times better in X ray, UV, optical, and near-IR than currently available are required for imaging faint, diffuse emission and resolving dense clumps over areas >300 kpc around Milky Way-type galaxies. High-throughput UV-optical spectrographs would enable absorption spectroscopy using QSO and galaxy background sources (see Question D-Q2b), sampling the diffuse CGM/ICM with a density of ~25 $arcmin^{-2}$, or an average grid spacing of ~100 physical kpc. Together, these capabilities would enable observations that constrain the density, ionization, metallicity, and velocity field of the CGM/ICM.

D-Q3. HOW DO SUPERMASSIVE BLACK HOLES FORM AND HOW IS THEIR GROWTH COUPLED TO THE EVOLUTION OF THEIR HOST GALAXIES?

Investigations of SMBHs over the past ~20 years have underscored their importance for galaxy evolution but also the significant gaps that persist in our understanding of these objects. We still do not know how SMBHs form and grow, how they interact with and impact their host galaxy and the CGM/IGM, the full range of BH properties (e.g., the shape, form and evolution of the BH mass function), and what role early BHs play in reionizing the universe. The coming decade promises answers to many of these questions.

D-Q3a. The Seeds of Supermassive Black Holes

The existence of luminous quasars at $z > 7$ requires SMBHs to grow to M ~ $10^9\ M_\odot$ in the challengingly short period, <1 Gyr, available since the Big Bang. Theorists have pursued a variety of ideas, broadly distinguished between "heavy seed" and "light seed" models. Examples of the former include the runaway collapse of early, ultra-dense stellar clusters, and the direct collapse of a primordial gas cloud into a >1000 $M_\odot$ BH, potentially as massive as $10^5\ M_\odot$. Alternatively, distant quasars could grow from light seed BHs, such as those formed from the death of massive stars, either through suppression of feedback that modulates inflowing gas accretion rates (i.e., super-Eddington accretion) at early cosmic epochs, or through rapid merging of stellar-mass BHs accompanying hierarchical structure formation at early times. Besides probing the origins of the most distant BHs, understanding the birth of BHs will inform us about early heating of the IGM, and teach us about a potential major source of feedback in primordial and low-mass galaxies.

Both heavy and light seed models have theoretical challenges, and observations will be required to discriminate among them. Proposed observational tests include (1) measuring the high-redshift ($z > 6$) quasar luminosity function (see Question D-Q1a for numbers and depth); (2) studying the occupation fraction of massive BHs in nearby low-mass galaxies; (3) detecting BH mergers down to $10^3\ M_\odot$ at $z \sim 10$ and ~a few × $10^3\ M_\odot$ at $z \sim 20$ using LISA; and (4) detecting high-redshift, massive seeds in the X rays. An actively accreting 10,000 $M_\odot$ BH at $z \sim 10$ requires arcsecond or better X-ray spatial resolution to avoid source confusion and enable unique host galaxy identifications, and will need to have that resolution at sensitivities $<10^{-19}$ erg/cm^2/s and over solid angles > 1 deg^2, the combination of which are far beyond Chandra's capabilities.

D-Q3b. Existence and Formation of Intermediate Mass Black Holes

There must be IMBHs in the gap between stellar origin BHs ($\lesssim 100\ M_\odot$) and SMBHs ($>10^5\ M_\odot$), as all viable paths to make SMBHs require a stage of "intermediate" mass. While IMBHs at the high end of the stellar mass range are now beginning to be observed, no BHs have yet been confirmed in the mass range 10^3–$10^5\ M_\odot$ that bridges the stellar mass regime with the BHs in the center of spheroids. Finding a population of IMBHs would be transformative, providing an evolutionary link in the growth of SMBHs and constraining formation channels of seed BHs in the early universe. The next decade will be ripe for discovery of IMBHs. Time domain surveys at X-ray, UV, and optical wavelengths will identify rare white dwarf tidal disruption events (TDEs), which probe BH masses $<10^5\ M_\odot$. Advanced LIGO may detect IMBHs with hundreds of solar masses, and LISA is expected to find merging 10^3–$10^5\ M_\odot$ BHs at $1 < z < 20$. IMBHs in local dwarf galaxies will be probed using several techniques, including LISA for detection of mergers with IMBHs and sensitive, high-angular-resolution UV, optical, and near-IR telescopes for resolving the gravitational sphere of influence of IMBHs (0.01" for a $10^4\ M_\odot$ BH at 5 Mpc; a few arcsec for Milky Way GCs) for both kinematic and integrated light studies. Measurements of integrated light profiles and proper motions will require tens of μ-arcsec relative accuracy, with samples of hundreds of

stars within the inner arcsecond of GCs for proper motions. Next-generation radio interferometers and X-ray observatories may detect radiative signatures of IMBH accretion. For X rays, high source sensitivity ($<10^{-19}$ erg/s/cm^2) at sub-arcsecond angular resolution in the ~0.5–10 keV band and, in the radio, sensitive imaging capable of resolving a few pc is needed to, for example, prove the existence of and locate IMBHs within local galaxies. Regardless of technique, key measurements include constraining the BH mass spectrum across its full range and measuring the fraction of halos harboring BHs as a function of mass. Such measurements will need to be coupled with advances in theoretical modeling of the underlying BH population and their hosts.

D-Q3c. Comprehensive Census of Supermassive Black Hole Growth

Recent NASA missions have made great progress in understanding the demographics of BH growth. The cosmic X-ray background, which is dominated by accreting BHs, is ~90 percent resolved into discrete sources by the deepest Chandra and XMM-Newton surveys at <6 keV, although this fraction falls at higher energies. At 8–24 keV, which overlaps with the 20–40 keV peak of the cosmic X-ray background, only ~35 percent of the background is resolved by the deepest NuSTAR surveys. While a large population of obscured AGNs exist, how this population depends on redshift, luminosity, source of the obscuring material (e.g., torus versus galactic dust and gas), and environment remain open questions. Mid-IR missions—for example, Spitzer and WISE—have detected obscured sources in large numbers, although they are biased to the high-luminosity AGNs where accretion luminosity dominates over stellar emission. The cosmic census of AGNs is currently patchy, which limits our understanding of the co-evolution of galaxies and their SMBHs. The highest redshifts remain largely unprobed, our knowledge of the most heavily obscured AGNs is incomplete, even at the lowest redshifts, and nuclear activity in the lowest-mass galaxies is poorly constrained. We need to fill these fundamental holes in our knowledge of AGN demographics in order to understand the mechanisms and importance of BH feedback in galaxy evolution, its interplay with star formation and star formation feedback, and address deep questions about the actual physical structure of AGNs.

JWST will probe obscured AGNs by studying polycyclic aromatic hydrocarbons (PAHs; rest-frame 3–9 μm), ionized neon (12–16 μm), and silicate absorption (~10 and 18 μm) out to $z > 5$ for the bluest features. New, sensitive mid-IR/far-IR spectroscopic capabilities will be required for the longer wavelength diagnostic features beyond $z \sim 2$. BH growth, particularly obscured BH growth, is likely enhanced during BH mergers, and our understanding of these mergers will grow with gravitational wave measurements by LISA and ground-based pulsar timing arrays, as well as with extremely high-resolution imaging by VLBI and future optical/near-IR facilities. Because BHs are multimessenger sources, progress will be enabled by many facilities, including deep X-ray surveys (e.g., Athena and hard X-ray, 10–30 keV, imaging surveys more sensitive than NuSTAR), deep optical/infrared surveys (e.g., Rubin Telescope, Euclid, and Roman), mid-IR observations (e.g., JWST), and time-domain surveys (e.g., Rubin Telescope).

D-Q3d. The Physics of Black Hole Feedback

The energy released by accreting SMBHs contributes feedback that helps regulate the growth of galaxies, although the magnitude and importance of that feedback is currently highly uncertain and likely varies for galaxies of different masses, environments, and evolutionary stage. Massive galaxies, groups, and clusters, which host the most massive BHs, provide a particularly promising way to distinguish proposed physical mechanisms (e.g., thermal, radiative, and/or magnetic processes) that accelerate outflows, because they allow study not only of the BH and its outflow, but also its impact on the surroundings. High spectral resolution X-ray kinematic measurements are a frontier scientific measurement for this question. Hitomi showed that the intracluster plasma of the Perseus cluster is

astoundingly calm, with a turbulent gas velocity of ~164 km/s. Some yet-to-be-understood process is suppressing energy pumped into the gas by the nuclear activity at the cluster core and prevents it from inducing turbulence into its surroundings. Regrettably, Hitomi was lost after only a few weeks of observations, so the immense promise of high-resolution nondispersive spectroscopy must now wait for XRISM and Athena. To probe AGN winds across all the relevant ionization states and phases, high-throughput, high-resolution spectroscopy from the hard X rays through the FUV is needed. Ultimately, arcsecond angular resolution in the X ray with much larger throughput than currently available (i.e., >1 m^2) is required for imaging—for example, shocks induced by outflows. Deep, spatially resolved infrared and millimeter measurements are required to probe the molecular outflows and outflow dust content. High-sensitivity radio interferometers in the 0.1–115 GHz range that reach below $L_{1.4\,GHz} \sim 10^{24}$ W/Hz with ≤100 pc resolution would probe and resolve jet-gas interactions out to $z \sim 1$, and enable precision studies of the ISM in both molecular (low-J CO in emission) and atomic gas (HI in absorption) to learn how such interactions occur, as well as to study the mechanical effects of BH feedback via study of radio bubbles.

D-Q4: HOW DO THE HISTORIES OF GALAXIES AND THEIR DARK MATTER HALOS SHAPE THEIR OBSERVABLE PROPERTIES?

Observational, semi-empirical, computational, and theoretical studies of the past decade have sharpened our understanding of the relation between galaxies and their host dark matter halos. Many aspects of this picture, and the progress achieved over the past ~10–20 years, have been summarized earlier in the “State of the Field”; yet, many aspects have not been strongly tested by observations, and the discovery potential is large for many areas in the field of galaxy formation over the coming decade. In particular, a plethora of data will parse our own Milky Way into its elemental constituents, enabling us to understand its physics and how singular or general it is as a system; in-roads in the investigation of the lowest-mass galaxies in the local universe will clarify the process of galaxy formation close to its mass threshold; and the characterization of the physical components of low-redshift galaxies will provide the benchmark for describing and understanding the galaxies at higher redshifts, where we do not have the luxury of high spatial resolution.

D-Q4a. The Dynamical and Chemical History of the Milky Way

The Milky Way affords unique insights into the governing processes of galaxy formation: although it is a singular example, we can study it at a level of detail impossible for other galaxies. The goal of observational and theoretical studies of the Milky Way is to understand the assembly, star formation, and chemical enrichment histories of the thin disk, thick disk, bulge, bar, and stellar halo; the origin of the striking bimodality of element abundance ratios across the disk; the importance of gas accretion, radial gas flows, fountains, and outflows through time and at the present day; the impact of dynamical perturbations on kinematic structure; the baryon content and temperature-density structure of the gaseous halo; and the mass, density profile, shape, and substructure of the dark matter halo. The final data releases from ongoing spectroscopic surveys and from the Rubin Telescope and Roman in the coming decade will greatly advance our knowledge of the structure and substructure of the Milky Way’s stellar components.

A benchmark goal for the 2020s is to increase the numbers of Milky Way stars observed at medium and high resolution by an order of magnitude (to $\sim 10^8$ and $\sim 10^7$ stars for $\lambda/\delta\lambda \sim 2000$ and $\lambda/\delta\lambda >$ 20,000, respectively) relative to current surveys, both by moving from grids of pencil beams to contiguous sky coverage *and* by reaching fainter spectroscopic targets. Contiguous coverage with multi-element spectroscopy, greater depth, and large numbers will decode the history of the disk and bulge, untangle the structure and merger history of the stellar halo, and measure perturbations to tidal streams

that could reveal the impact of dark matter substructure. The thousands of very metal-poor stars ([Fe/H] < –2) secured by these new surveys will probe chemical enrichment during the earliest phases of the Milky Way's formation. Taking full advantage of these data advances will require corresponding advances in calibration methods for abundance measurements, in statistical methods for interpreting enormous high-dimensional data sets, and in numerical simulations that resolve the intricate details of the formation of galaxies like the Milky Way.

At the same time, better characterization of the gaseous components of the Milky Way is essential for understanding the physics that governs the Milky Way today and for providing a template to interpret observations of other galaxies. Higher sensitivity X-ray spectroscopy can greatly expand the set of AGN sightlines that can probe hot gas absorption in the halo, provided that the absorption lines can be separated from those of the Milky Way ISM and Local Group galaxies, while improved wide-field X-ray IFUs can provide emission-line detections. Simultaneous measurements of OVII and OVIII in absorption and emission provide strong diagnostics of density and temperature structure. In the Galactic Center, progress is still required to understand the extreme environments of the ~2 pc Circumnuclear Ring and the ~200 pc Central Molecular Zone—their energetics, motions, physical characteristics, and so on—and to provide comparison templates for more distant galaxies. The high dust content of this region requires use of infrared and longer wavelength probes, like those provided by SOFIA and (sub)mm/radio facilities.

D-Q4b. The Threshold of Galaxy Formation

The comparison between observed numbers of galaxies and predicted numbers of dark matter halos implies that the efficiency of galaxy formation plummets in halos with mass below M_{halo}~10^9 $M_\odot$, and observable stellar systems do not form in halos with mass $M_{halo}<10^8$ $M_\odot$. The lowest-mass galaxies, M_{star} ~10^3–10^7 $M_\odot$ that inhabit halos close to this threshold are unique crucibles for galaxy formation theory. These are the most dark-matter-dominated and chemically primitive galaxies in the universe, and their shallow gravitational potential wells render their baryonic content sensitive to a variety of feedback processes. Open questions include the number of surviving satellite galaxies residing within halos, the number of destroyed low-mass galaxies that populate the stellar halos of their central galaxy hosts, and whether ultra-faint galaxies are fossils of the reionization era or can form in substantial numbers at lower redshifts. These systems have been studied in detail only in the Local Group or, for the lowest-mass ultra-faint galaxies that now blur the boundary with GCs, within the inner tens of kpc of the Milky Way's dark matter halo.

An area ripe for major advances is pushing to larger distances and to different environments. Initial results on the M31 satellite system have revealed intriguing differences in its population relative to that of the Milky Way, and the next decade will see dramatic improvements in our understanding of ultra-faint galaxies beyond the Milky Way. In particular, the Rubin Telescope sensitivity limit over 10 years of operations will reach 2000 $L_\odot$ galaxies out to 1 Mpc from the Milky Way (~3–4× the virial radius) and galaxies as faint as 200 $L_\odot$ throughout the virial volume of the Milky Way; its lifetime sensitivity will detect classical dwarf galaxies (L_V ~ 10^5 $L_\odot$) with half-light radii >1 kpc. This will open up the possibility of studying, in detail, ultra-faint galaxies that have never interacted with the Milky Way or M31—galaxies that have never been discovered but must exist if current CDM-based models of galaxy formation are correct—thereby providing important information about the effect of environment, reionization quenching, and the halo-galaxy connection at the lowest possible masses. JWST will enable observations of old main sequence turn-off stars (M_V = +4) to characterize star formation histories at the earliest epochs out to ~3 Mpc. Large fields, like those covered by the Rubin Telescope and Roman, will enable efficient observations of typical Milky Way dwarfs, with high target density for simultaneous observation of nearby stars. Wide-field, high-resolution, multi-object spectroscopy will bring progress with detection of weak and narrow metal lines in the oldest Pop II stars ([Fe/H] < –3) and velocity accuracy of 1 km/s to characterize the dynamics of the lowest mass systems (σ ~ 4–5 km/s) and resolve stellar binary motions.

D-Q4c. Connecting Local Galaxies to High-Redshift Galaxies

Nearby galaxies serve as anchors for our interpretation of the physical and chemical histories of individual galaxies and galaxy populations at earlier epochs, including before, during, and soon after the EoR. Two of the key challenges for interpreting high-redshift observations are determining which galaxies produce and leak LyC photons and calibrating indicators of star formation, metallicity, and dust content.

Production of LyC photons is linked to the characteristics and evolution of the most massive stars, which are found in young, massive and supermassive ($M_{stars} > 10^4\ M_\odot$) star clusters. Collecting statistically significant numbers of these relatively rare clusters requires probing crowded regions within galaxies in the local ~50–100 Mpc, using unique UV spectral signatures (e.g., P-Cygni NV and CIV, and broad HeII) to characterize their ionizing stellar content. Establishing the conditions, internal and/or external, under which galaxies leak LyC photons and calibrating the indirect UV and optical diagnostics to be used at $z > 4$ will require mapping of LyC photon leakage from galaxies at $0.1 < z < 3$ in sufficiently large numbers to discriminate among different conditions for escape.

Calibrating abundance measurements with UV nebular lines will be a top priority for the interpretation of the chemical build-up of galaxies at $z > 8$ with JWST. Reconciling the metallicity scales of nebular emission lines, neutral gas absorption lines, and stellar photospheric lines will enable the interpretation of the chemical history, transport, and mixing, and the ionization structure of galaxies across cosmic times. These will require tracing the faint UV (HeII, CIII], OIII], SIII], etc.) and optical (HeI, auroral lines) lines within and across HII regions, and measuring abundances and depletion patterns of key elements in the neutral gas and in the photospheres of stars across the full range of metal abundances in nearby galaxies.

Metallicity calibrations and the quantification of LyC production and escape will require various combinations of sensitive, wide-field, high-spatial (~a few pc at 100 Mpc) and low-to-high spectral (from $\delta v \sim 500$ km/s to $\delta v \sim$ a few km/s) resolutions, UV and optical IFUs and multi-object spectrographs capable of detecting and characterizing faint photospheric lines and gas emission and absorption in local to medium-redshift galaxies. Recent large optical IFU and wide-field spectroscopic surveys (e.g., CALIFA, MaNGA, SAMI) offer a roadmap for how the approach can be extended to the UV. Progress in the theory and modeling of massive star properties and evolution, including the role of multiplicity, rotation, and so on, will need to accompany observations in order to enable their interpretation.

Quantifying the SFR of high-redshift, dusty galaxies requires use of infrared tracers, such as the [CII] fine structure line. An accurate calibration of this tracer, both in intensity and line shape, as a function of local environment in the Milky Way and nearby galaxies can be accomplished with SOFIA, building on the legacy of Herschel.

D-Q4d. The Evolution of Morphologies, Gas Content, Kinematics, and Chemical Properties of Galaxies

Kinematics, metal abundances, and gas content of galaxies are unique tracers that connect evolving galaxy populations across time and help reveal how galaxies obtain their present-day structures. The past decade has provided us with the first measurements of the kinematics and resolved chemical abundances in galaxies at $z = 1$–3, showing well-defined trends but also more diversity than in the local universe. Observational capabilities have restricted such studies to $M_{stars} \gtrsim 10^{9.5}\ M_\odot$, with limited spatial information. As a result, we have essentially no detailed information about progenitors of typical disk galaxies like the Milky Way at the peak of cosmic star formation rate density ($z \sim 2$) or earlier; this is also a major limitation in extrapolating the detailed observations being made in the Milky Way to the full population of similar galaxies at all cosmic epochs. JWST will begin to provide detailed kinematic and metal abundance maps at higher redshifts and lower masses and will reveal how galaxies transition from disordered kinematical and morphological states to more ordered systems (e.g., disks). The gain in

sensitivity and wavelength coverage of JWST relative to current facilities will open up the high-redshift regime, but its limited spatial resolution (~0.1"~0.7 kpc at $z \sim 4$–7) will resolve only the largest, most massive galaxies. In addition, absorption-line-based work for spatially resolved stellar kinematics and metallicities will remain extremely challenging with JWST owing to both sensitivity and resolution limitations. Further progress in this area will require multi-object near-IR spectroscopy with HII-region-size resolution and $\delta v \sim 50$ km/s, with enough sensitivity to map galaxies in the stellar continuum below the knee of the stellar mass function ($M_{stars} \sim 10^{10}\ M_\odot$) from $z \sim 1$ to the end of reionization at $z \sim 6$. An important benchmark for testing the paradigm established over the past decade in which bursty star formation gravitationally heats the central regions of the host dark matter halos is to reach $M_{stars} \ll 10^9\ M_\odot$ systems, where this feedback-induced effect is predicted to be most efficient. Deep ALMA observations of the restframe far-IR [CII] line for large samples of galaxies out to z ~ 7–8 will secure the full census of the star formation in the dusty progenitors of today's massive galaxies. Bulk metal abundance measurements in these obscured galaxies will require temperature-insensitive tracers—for example, [OIII] in the far-IR, that cannot be observed from the ground below $z \sim 7$.

DISCOVERY AREA: MAPPING THE CIRCUMGALACTIC MEDIUM AND THE INTERGALACTIC MEDIUM IN EMISSION

Imaging the CGM and IGM in emission out to $z \sim 10$ and beyond is a major opportunity for the next and the following decade, one that is just now coming into view with new instruments and observing strategies. "Imaging" is intended as contiguous sky coverage, over fields large enough to probe significant volumes around and between galaxies. This implies fields-of-view (or contiguous mapping) of several arcminutes or more, subtending ≳1 Mpc at $z = 2$–10. The goal is to detect line emission, from neutral hydrogen to ionized gas and metals across multiple spatial scales, which requires spatially resolved spectroscopy probing from the kpc-size distribution of diffuse HI down to the ~100 pc sizes of HII regions within galaxies. These observations will address or contribute to addressing the science questions in this appendix, including (1) a full baryon and metal accounting in different gas phases at different redshifts, (2) how galaxies acquire fuel for sustaining star formation, and (3) a high-fidelity image of how energy and momentum from stars and SMBHs are transferred to the low-density CGM/IGM as a function of time (see topics D-Q1a, D-Q2a, D-Q2b, D-Q2c, D-Q2d, D-Q3d, D-Q4a, D-Q4c, D-Q4d). Thus, it is worth exploring innovative designs of instruments for integral field emission-line mapping that accommodate needs for a range of spatial resolutions. Degree-size, low-angular resolution spectral maps offer complementary information to smaller-field, high-angular resolution spectral maps, by capturing the integrated emission from galaxies too faint to be detected individually and by averaging over cosmic variance, which enables spatial cross-correlations of multiple tracers to constrain physical processes.

We have learned an impressive amount about the IGM and CGM from absorption-line observations, in part because cosmological simulations have proven effective at creating a synthesized picture from the individual sightlines. However, simulation predictions are sensitive to numerical uncertainties and to unconstrained physics, including the geometry, kinematics, enrichment, and physical conditions of galactic winds, the impact of thermal instability, mixing instabilities, metal diffusion, and conduction, and the interactions between outflows and accretion. Emission-line maps across all spatial scales linking the galaxy, through the CGM, to the IGM can test all aspects of these predictions and ultimately provide observations of the cosmic baryon distribution and circulation that rival the level of detail that we currently have only from theory (see Figure D.1). This will place constraints on the physics of the mechanisms that regulate galaxies and larger structures, and drive their evolution.

The restframe UV-optical-IR provides a rich suite of emission lines, including HI and He recombination lines, together with strong metal lines, SiIV, CIV, OVI, and lower ionization metal species, such as OII, OIII, SiIII, and CIII, which typically trace cooler gas. These lines can be accessed from either space or the ground, depending on redshift. Wide-field IFU spectroscopy, with spectral resolution of a few thousand, affording velocity resolution $\delta v \sim 50$ km/s, is required to match typical

velocity spreads in halos, and measure gas kinematics. The ISM in galaxies is clumpy over ~100 pc sizes, and metal line emission from shock-heated gas in the CGM and IGM is also expected to be clumpy, requiring ~0.01"–0.1" imaging to resolve, while diffuse HI will benefit from lower angular resolution. JWST will offer sensitive multi-object spectroscopy, but without contiguous spatial coverage, or small-field IFU capabilities. SPHEREx will provide groundbreaking all-sky coverage at 0.7–5.0 μm, but at low resolution, both spectral (R ~ 100) and spatial (6"–10"). At least an order-of-magnitude increase in both spatial resolution and spatial coverage relative to JWST IFUs are required for CGM/IGM imaging that informs theory. The greatest challenge, however, is sensitivity. Fluorescent Lyα emission from optically thick HI ($N_{HI} \sim 10^{18}$ cm^{-2}) illuminated by the metagalactic UV background has a predicted line surface brightness of ~ 10^{-20} erg s^{-1} cm^{-2} arcsec^{-2}, although this is higher in the vicinity of bright quasars where the UV background is boosted. Simulation predictions for metal-line emission are uncertain but detecting bright features in the outer regions of halos (r ~ 100 kpc) also requires sensitivity ~10^{-20} erg s^{-1} cm^{-2} arcsec^{-2} or better. This is deeper by up to an order of magnitude than what is routinely achievable on 10 m class telescopes. Thus, mapping the intricate morphologies visible in simulations of high-redshift structures will require IFUs on much larger telescopes than are currently available.

Moving to hotter gas, the E/dE ~ 100–1,000 resolution at < 1 keV and 3 arcmin field of the XRISM micro-calorimeter will allow emission mapping of OVII, OVIII, and Fe lines (FeXXV and FeXXVI) in the ICM, and in the gas at the outskirts of local groups and clusters out to z ~ 0.3. The larger effective area of Athena allows maps of continuum and line emission to the virial boundary and beyond for clusters and groups, while also providing sensitivity to the hot gas halos of individual L^* galaxies. Ultimately, still larger effective areas over large field-of-views, of order of tens of arcmin, and high spectral resolution are needed to reach the shock-heated phases of the diffuse IGM. High spectral resolution (E/dE ~ 2000) and ≲1" angular resolution are necessary for resolving the multiphase structure of the ICM/CGM and the discrete sources that contribute to the cosmic X-ray background.

Sunyaev-Zel’dovich (SZ) distortions of the cosmic microwave background (CMB) provide a complementary way to map ionized gas. The combination of X-ray and SZ maps is a much stronger diagnostic of density and temperature structure than either observable on its own. Next-decade facilities will push the limits to a few times higher angular resolution and higher sensitivity than the Planck satellite, with SZ detection thresholds of individual halos down to ~10^{14} $M_\odot$ over large areas of sky, extending to 10^{11}–10^{12} $M_\odot$ (~10–100× deeper than current limits) for stacking analyses. These will provide new constraints on cumulative energy injection into the CGM. The combination of kinetic SZ measurements from these CMB experiments with new galaxy redshift surveys (e.g., DESI, SPHEREx, and Euclid) will enable cross-correlation of the CMB signal with the peculiar velocity field of galaxies and provide novel insights into the electron density distribution across many different circumgalactic and cosmological environments. New and more powerful CMB experiments will offer a rich new probe of the cosmic baryon distribution by extending to the denser and hotter phases of the IGM beyond halo virial radii.

Large, contiguous maps of cool gas (≲10^4 K) complete the picture above. Existing maps of atomic hydrogen and molecular gas mostly probe the ISM of star-forming galaxies. However, observations of galactic winds show that they frequently contain atomic and molecular gas; whether this is entrained from the ISM or cools out of the hot flow is unclear. High-sensitivity observations over large fields could detect neutral hydrogen in the more distant CGM, where Lyman Limit absorption ($N_{HI} > 10^{17}$ cm^{-2}) is frequently observed, and molecular gas if it is present. Direct detection of HI or CO emission from the CGM, at distances of tens or even hundreds of kpc, would be a powerful diagnostic of thermal instability in the CGM and the geometric interleaving of cold gas accretion and hot gas outflows. Before reionization, most neutral hydrogen resided in the IGM rather than in galaxies. Mapping the z > 8 era in the redshifted 21 cm line may lie beyond the capabilities realized in the 2020s, but such maps will eventually provide an extraordinary view of intergalactic gas at an epoch when few galaxies existed. The needs and expected progress in this area are discussed in Appendix C.

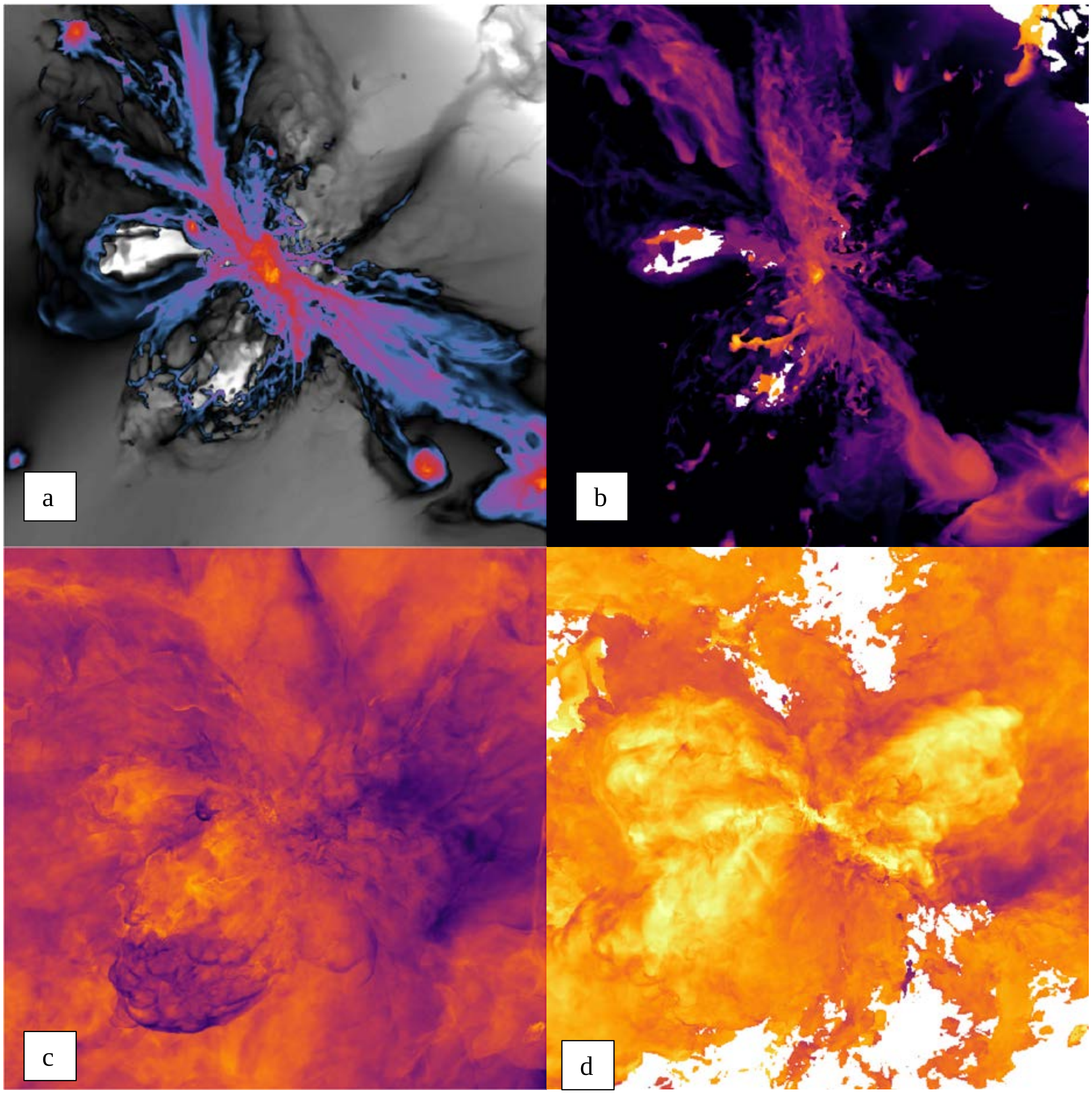


FIGURE D.1 The circumgalactic and intergalactic medium around a simulated Milky Way progenitor at $z = 2$. At this redshift, the mass of the galaxy's host halo is 3.9×10^{10} M_{sun}. The panels are 200 h^{-1} kpc (comoving) on a side, subtending an angle of 11 arcsec at $z = 2$. *(From upper left)* The column densities of (a) neutral hydrogen and of (b) the total metals in cool gas ($T < 10^5$ K), (c) in warm gas (10^5 K < T < 10^6 K), and (d) in hot gas (10^6 K < T < 10^7 K). Emission-line maps that trace these components can reveal filamentary accretion and bipolar outflows and test predictions of the metallicity, thermal, and velocity structure of the CGM. SOURCE: Figure courtesy of M. Peeples, 20202; based on Peeples et al., 2019, *ApJ*, 873, 2.

SUMMARY AND FINAL CONSIDERATIONS

The past two decades have firmly established the ΛCDM framework that sets the initial conditions for cosmological structure formation and for galaxy formation and evolution within and connected to these large-scale structures. In the coming decade, a suite of powerful facilities with unprecedented capabilities for studying galaxies will begin operations, priming the field of galaxy formation and evolution for a period of major advances.

The four questions and the discovery area outlined in this appendix represent key areas of exploration. The appendix identifies observations, theory investigations, and simulations necessary to either advance our knowledge or fully answer each of those questions in the next decade, and to set the stage for the decade that follows. The progress required to address those questions will also greatly advance many other areas of research not explicitly mentioned in this appendix. A common theme throughout the appendix is the need for a synergistic multiwavelength approach and for continuous interaction between observations and theory. Theory and simulations will need to progress toward connecting multiple physical scales across many orders of magnitude in dynamical range: from detailed stellar models to stellar population synthesis, galaxy models, zoom-in galaxy simulations, and large box simulations of the universe. Curation of multimission/facility, cross-referenced data archives, quickly searchable across multiple dimensions and parameters, will also be required for the success of the program outlined in this report. The huge, petabyte-size data sets that will become available will require novel approaches to data handling and new statistical methods for data analysis and interpretation. Data volume represents a true challenge for the next decade, which will require innovative approaches in order to fully realize its promise.

A summary table (Table D.1) showing the flow from the science questions and topics to the observational and theoretical needs as presented in this report is given below.

TABLE D.1 Summary Table

Science Question	Sub-Topics	Future Needs
DQ-1. How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?	DQ-1a. Detailed thermal history of the intergalactic medium and the topology of reionization. DQ-1b. Production of ionizing photons and their escape into the intergalactic medium. DQ-1c. Properties of the first stars, galaxies, and black holes.	• Wide-field NIR imaging of $>10^5$ $z \sim 10$–12 $M_{star} \sim 10^8\, M_\odot$ galaxies, and of hundreds of $z \sim 15$–20 $M_{star} \sim 10^{6–7}\, M_\odot$ galaxies with ~100 pc resolution. • Wide-field NIR imaging to AB $\gtrsim$ 35 for direct detection of Pop III stars; degree-size fields to AB ~ 31 for indirect Pop III stars detection via pair instability supernovae. • NIR multi-object spectroscopy, with $\delta v \sim 50$–100 km/s and ~100 pc resolution, to characterize hundreds of $z > 8$ $M_{star} \sim 10^{5–7}\, M_\odot$ galaxies; NIR single-object spectroscopy with $\delta v < 10$ km/s of $z = 7$–9 QSO proximity zones. • Mapping of HI 21 cm at $z \sim 6$–12, more than tens of deg^2. • Alert system for very high-z GRB follow-up. • Theory/simulations: models of formation and evolution of first stars and galaxies; physics of IGM reionization.
DQ-2. How do gas, metals, and dust flow into, through, and out of galaxies?	DQ-2a. The acquisition of the gas necessary to fuel star formation. DQ-2b. The production, distribution, and cycling of metals.	• Wide-field X-ray (0.3–10 keV), UV (0.09–0.3 μm), and optical/IR (0.3–2.5 μm) IFU spectroscopy, $\delta v < 100$ km/s, to map warm/hot gas (T ~ 10^4–10^7

	DQ-2c. The coupling of small-scale energetic feedback processes to the larger gaseous reservoir. DQ-2d. The physical conditions of the circumgalactic medium.	K) and characterize gas and metals in/around galaxies, over contiguous ~0.01–1.0 Mpc at $z \geq 1$ and down to ~100 pc resolution at $z \sim 4$. • Multi-object UV/optical absorption spectroscopy, $\delta v <$ 10 km/s, for faint ($UV_{AB} \sim 23$) background QSOs and galaxies. • Cold molecular gas in $10^8\ M_\odot$ sub-kpc clumps with $\delta v \sim$ 10–30 km/s at $z > 2$ in typical star-forming galaxy. • Theory/simulations: physics of feedback, role of magnetic fields and cosmic rays, and IGM/ICM/CGM/galaxy/star connections across spatial scales.
DQ-3. How do supermassive black holes form and how is their growth coupled to the evolution of their host galaxies?	DQ-3a. The seeds of supermassive black holes. DQ-3b. Existence and formation of intermediate mass black holes. DQ-3c. Comprehensive census of supermassive black hole growth. DQ-3d. The physics of black hole feedback.	• Multi-λ time domain surveys for TDEs. • MHz gravitational waves, to detect ~$10^3\ M_\odot$ BHs at $z < 20$; pulsar timing arrays to $10^9\ M_\odot$. • X-ray (0.5–2 keV) imaging with sufficient field-of-view and sub-arcsec resolution to detect $10^4\ M_\odot$ BHs at $z \sim 10$. • Wide-field hard X-ray (10–30 keV) imaging and spectroscopy for SMBH census and shock-induced outflows. • Wide-field X-ray (0.3–10 keV) and UV (0.09–0.3 μm) IFU spectroscopy for hot gas and feedback physics. • Optical/NIR (0.3–2.5 μm) μ–arcsec precision astrometry and 0.01" resolution IFU spectroscopy for BH masses from stellar proper motions and kinematics. • Mid-IR/far-IR (~30–500 μm) imaging and spectroscopy for obscured QSO census and diagnostics at $z > 2$. • Radio, 0.1–115 GHz for 100 pc scale HI in absorption and CO from $z < 1$ jet-gas interactions; 1–90 GHz imaging at sub-kpc resolution for IMBHs and synchrotron emission from AGN jets at $z < 1$.

		• Theory/simulations: BH seeds formation and evolution, BH mass function.
DQ-4. How do the histories of galaxies and their dark matter halos shape their observable properties?	DQ-4a. The dynamical and chemical history of the Milky Way. DQ-4b. The threshold of galaxy formation. DQ-4c. Connecting local galaxies to higher redshift galaxies. DQ-4d. The evolution of morphologies, gas content, kinematics, and chemical properties of galaxies.	• All-sky optical/NIR multi-object spectroscopy for abundances and kinematics of ~10^8 MW stars and stars in satellite dwarf galaxies. • Wide-field, sub-arcsec resolution, X-ray IFU spectroscopy for MW halo; UV IFU spectroscopy for metallicity calibrations and LyC measurements. • Wide-field, optical/NIR multi-object spectroscopy at ~100 pc resolution to map galaxies to $z \sim 6$. UV multi-object spectroscopy at 3–5 pc resolution for massive star properties and metallicity calibrations in nearby galaxies. • Mid-IR/far-IR (~30–500 μm) spectroscopy for bulk metallicities to $z \sim 2$. • Theory/simulations: next-generation numerical simulations of Milky Way-like galaxies formation. Modeling of properties and evolution of massive stars in stellar populations.
Discovery Area: Mapping the circumgalactic medium and intergalactic medium in emission.		• Wide maps of neutral and ionized gas emission lines of galaxies/CGM/ICM/IGM, both intensity and kinematics, out to ~0.3–1 Mpc radius, $\delta v \sim 50$ km/s, at X-ray, UV, optical, IR, radio (including HI 21cm), with resolution from ~100 pc to $\lesssim$1 kpc. • Next-generation SZ experiments down to $M_{halo} \sim 10^{14}\ M_\odot$ for individual halos over large areas of sky.

E

Report of the Panel on Exoplanets, Astrobiology, and the Solar System

OVERVIEW

In the past decade, the field of exoplanet science has rapidly expanded with the discoveries of thousands of new planets, and the characterization of worlds unlike those in our solar system. From the ensuing treasure trove of exoplanet demographics and characteristics, we have learned that most, if not all, stars host planets, and that planets smaller than Neptune are ubiquitous. These systems and the planets that comprise them are surprisingly diverse, with few matching the solar system. We have characterized a plethora of larger worlds for density and atmospheric properties, progressing from gas giants to large terrestrial planets, and from highly irradiated planets to cooler planets, as observational sensitivity and techniques improved. Significant advances have been made in understanding solar system planetary processes and how our planetary system formed and evolved. We have expanded our understanding of formation processes and the subsequent interactions of exoplanets with their host stars, and other components of their planetary systems, and identified planetary migration as a common process for exoplanet systems and our solar system. Complementing our studies of individual worlds, multiple techniques have pieced together a broad understanding of exoplanet classes, enabling a new era of comparative planetary system science as we work toward a more complete census.

Even though exciting progress has been made, significant key advances are still needed to place the solar system and our inhabited Earth in its cosmic context. Although we have discovered and characterized giant planets close to and very far from their stars, analogs of solar system giant planets have been beyond our reach. We have discovered close-in likely terrestrial[1] exoplanets, but none in the habitable zone[2] (HZ) of G dwarfs like our Sun. A handful of terrestrials are known to orbit in the HZ of M dwarfs, but we have not been able to probe their atmospheres to understand if they are truly Earth-like, or had strongly divergent evolutionary paths. We understand that interactions within the entire planetary system are critical to understanding the formation, evolution, environment, and habitability of exoplanets, but interdisciplinary research is still needed to better understand planets as interacting components evolving in the context of their host star and planetary system environment.

Upcoming observations of terrestrial exoplanets will enable one of humanity's grandest explorations—the search for habitable environments and life around a diversity of nearby stars. In the near term, the James Webb Space Telescope (JWST) and ground-based telescopes will have the sensitivity to search for and begin to characterize the atmospheres of a handful of terrestrial planets orbiting the closest M dwarf stars. Even more ambitious direct imaging missions will be needed to study habitable zone worlds orbiting more Sun-like stars. Exploring this exciting and unprecedented frontier will help place Earth's sparkling oasis of life in its cosmic context. This search is now within our scientific and technological reach, and can be informed by studies of larger exoplanets and solar system

[1] A terrestrial planet has a bulk composition dominated by rock and iron, such as Mercury, Venus, Earth, and Mars. "Terrestrial" does not imply that the planet is truly "Earth-like"—that is, habitable.

[2] The habitable zone is that region around a star where an Earth-like planet is considered more likely to be able to support surface liquid water.

analogs, as well as interdisciplinary efforts that incorporate theory and laboratory investigations. The next section outlines key discoveries in the past decade that set the stage for exciting future advances.

PROGRESS IN EXOPLANET, ASTROBIOLOGY, AND SOLAR SYSTEM SCIENCE SINCE *NEW WORLDS, NEW HORIZONS*

Exoplanet Detection and Planetary System Architectures

Since 2010, the number of known exoplanets has increased by an order of magnitude to more than 4000, with large contributions from both radial velocity (RV) and transit surveys. Early on, the typical planets detected were massive Jovians orbiting within a few astronomical units (AU) of their stars. The 2009 launch of Kepler inaugurated an era of thousands of discoveries, detecting significantly smaller, but still close-in, transiting planets (1–4 Earth radii). Most are closer than Mercury is to the Sun, but, because they orbit cooler stars, several are within their star's HZ. In parallel, microlensing surveys detected planets near M dwarfs snowlines, while direct imaging refined our view of the outer reaches of planetary systems.

The Distribution and Nature of Giant Planets

Over the 25 years since the Nobel Award-winning discovery of 51 Pegasi b, our understanding of giant planets has matured significantly. RV surveys have increased in sensitivity by orders of magnitude and observational campaigns begun in the 1990s now have the baselines required to detect giant planets with orbital periods similar to those of Jupiter. These surveys have revealed that hot Jupiters like 51 Peg b are rare, and that close-in brown dwarfs are rarer still. While hot Jupiters are not common, their frequency increases around more metal-rich stars, indicating that present-day system architectures are partially set by the initial mass and composition of the protoplanetary disk in which they form. Mass measurements of transiting planets have revealed a large range of planetary radii at a given mass (3 orders of magnitude in density), especially for planets near Neptune mass, suggesting a diversity of compositions even at fixed mass. Giant exoplanets are enriched in heavy elements compared to their parent stars, and this enrichment seems to increase with decreasing planet mass, mirroring the trend seen in solar system planets.

Although many planets have been found, our understanding of giant planets at a range of orbital distances comparable to those in our solar system is largely incomplete. RV and direct imaging surveys of young stars find that only 10 percent of solar-type stars harbor giant planets between 1 and 13 Jupiter masses inside of 100 AU, with such planets being more and less common, respectively, around higher and lower mass stars. Roughly a dozen planets are known with semimajor axes larger than 50 AU, and many have poorly constrained orbits and masses. Based on all available surveys, the occurrence rate of gas giants appears to peak near a few AU and then decline at larger separations, but these estimates depend on extrapolations of power laws in mass and semi-major axis, and the ~3–10 AU region is not yet thoroughly explored. A more complete census would be needed to determine if our solar system is unusual in having a Jupiter, which has large implications for planetary evolution and the delivery of "volatiles"—water and key compounds involving C, H, N, and O that condense at lower temperatures—which can be delivered to drier inner planets by more bodies that form farther out.

The Distribution and Nature of Sub-Neptune Planets

Three of Kepler's key discoveries were that sub-Neptunes (1–4 Earth radii) are the most abundant type of exoplanet at orbital periods <200 days, that ~50 percent of stars have small planets orbiting more closely than Mercury orbits the Sun, and that M dwarfs host close-in planets at a higher frequency than

Sun-like stars. Comparing the masses and radii of exoplanets to theoretical predictions and solar system planets has started to reveal key compositional trends. At orbital periods shorter than Mercury's, intense stellar radiation has sculpted the mass-radius diagram, inflating hot Jupiters and producing a bimodal radius distribution for small planets, likely owing to atmospheric escape. While highly irradiated planets smaller than ~1.6 R_{Earth} have bulk densities consistent with a terrestrial composition, larger planets require significant fractions of volatiles, as do solar system ice giants. Near-terrestrial masses and bulk densities have been measured for transiting planets in near-resonant configurations using transit timing variations (TTVs), and spectroscopy of white dwarfs possibly polluted by disrupted planets/planetesimals have revealed abundance ratios similar to those of the bulk Earth. However, our view of the mass-radius diagram is still dominated by planets larger and hotter than Earth. Tracking the existence and location of the planet radius gap as a function of stellar mass, stellar metallicity, and lower insolation will refine our view of the formation and evolution of low-mass planets, and will help determine whether some terrestrials are the evaporated cores of larger planets that have lost their natal volatile-rich atmospheres.

We have discovered ~20 likely terrestrial planets ($R < 1.6\ R_{Earth}$ and roughly terrestrial densities) within the HZ. Although many of these planets are too distant for follow-up characterization, Kepler's sensitivity has enabled more precise estimates of the frequency of potentially habitable planets in the Milky Way. The frequency of Earth-like planets orbiting Sun-like stars was unconstrained before the launch of Kepler, but detailed analyses of Kepler data have revealed that such planets are relatively common and occur on average around ~10 percent of Sun-like stars and ~20 percent around smaller, cooler red dwarf stars[3,4,5]). Ongoing ground- and space-based surveys of nearby cool dwarfs have discovered a small but growing number of HZ terrestrial planets with atmospheres accessible to JWST and potential future extremely large ground-based telescopes (ELTs).

Mapping Planetary Architectures: The Solar System in the Context of Exoplanetary Systems

Our understanding of planetary system architectures is currently in its infancy. While we have limited sensitivity to solar system-like planets, initial indications of the rarity of Jupiter analogs suggest that solar system-like architectures may be rare also. Transit measurements have discovered the most multiplanet systems to date, but are biased toward finding largely co-planar, close-in, tightly packed systems that would fit within Mercury's orbit, and show signs of migration. The RV method has found more widely spaced multiplanet systems, being sensitive to gas giants at longer orbital periods.

Growing knowledge of planetesimals distributions in our own debris disk (Kuiper Belt Objects [KBOs], comets, and asteroids) has modified our understanding of the solar system from an arrangement of stationary planets to a complex system of migrating planets. Models for the early migration of solar system planets, such as the Nice and Grand Tack models, reproduce many of the observed planetesimal distributions, providing a system-wide connection of solar system bodies, and predictions for exoplanet outcomes.

The distribution of material in mature debris disks can also inform the history of exoplanetary systems. Spatially resolved visible-NIR images of dozens of bright debris disks, analogous to more massive versions of our Kuiper Belt, show extended halos of dust in the cold outer regions, potentially sculpted by the interstellar medium (ISM). Atacama Large Millimeter/Submillimeter Array (ALMA)

---

[3] R. Belikov, et al., NASA's Exoplanet Exploration Program Analysis Group (ExoPAG), 2017, "SAG 13: Exoplanet Occurrence Rates and Distributions," https://exoplanets.nasa.gov/system/presentations/files/67_Belikov_SAG13_ExoPAG16_draft_v4.pdf.

[4] J.J. Fortney, T.D. Robinson, S. Domagal-Goldman, D.S. Amundsen, M. Brogi, M. Claire, M.S. Marley et al., 2016, The need for laboratory work to aid in the understanding of exoplanetary atmospheres, arXiv preprint arXiv:1602.06305.

[5] C. Dressing and D. Charbonneau, 2015, The occurrence of potentially habitable planets orbiting M dwarfs estimated from the full Kepler dataset and an empirical measurement of the detection sensitivity, *Astrophysical Journal* 807:45.

images of debris disks show underlying planetesimal distributions that are typically well-defined belts, indicative of sculpting by planets. Silicate emission from copious hot dust and density asymmetries in cold belts suggest possible collisional events, and disks that are more dynamic than previously thought. Directly imaged variations in the AU Mic disk resemble material being ejected by stellar winds. Dust compositions and optical properties are varied and poorly constrained, but likely silicate and water dominated. Detection of low levels of gas in debris disks via atomic absorption and molecular CO emission suggest nonsolar compositions, with significant carbon enhancement in some systems. The two populations of imaged debris disks and systems with known planets have little current overlap, in part owing to disk imaging sensitivity limits. Yet many disks include belts and inclination warps, likely owing to exoplanets. Earth has left its imprint on the solar system's disk by gravitationally shepherding zodiacal dust into a large, clumpy circumsolar ring; extrasolar planets should also create these telltale signposts of planets in debris disks, but so far these structures have eluded detection and may be limited to fainter disks currently below detection limits.

Exoplanet Characterization and Solar System Synergy

Efforts to characterize and model exoplanet atmospheres have focused largely on giant and Neptune-size planets; atmospheric characterization of smaller planets has just begun. Comprehensive surveys of transiting planets across a range of mass, radius, orbits, and/or insolation levels have provided key insights into interior and atmospheric composition, as well as the atmospheric circulation, chemical, and radiative properties that regulate planetary atmospheres. Observational studies have compared the atmospheric composition of dozens of planets. For directly imaged planets, spectroscopic and photometric observations have measured the abundances of multiple molecular species (H_2O, CH_4, CO) and revealed the presence of cloud decks, setting young giants on a continuum with more massive brown dwarfs. For transiting planets, atmospheric characterization first focused on more easily detectable atoms and molecules (Na, K, and H_2O) and expanded with improved observing methodologies and capabilities.

The physical conditions in planetary atmospheres, which probe processes like global circulation and radiative energy balance, have been thoroughly studied for roughly a dozen larger planets. High-resolution spectroscopy has measured precise thermal profiles, winds, and rotation rates for a handful of giant planets. The Hubble Space Telescope (HST) and Spitzer thermal phase curves have constrained atmospheric circulation by comparison to 3D general circulation models, and HST and ground-based high-resolution spectra have detected thermal inversions arising within strongly absorbing atmospheric regions. The same techniques have detected atmospheric escape from several hot, gas-rich transiting exoplanets, confirming that escape is common and may influence the size of close-in planets. Planetary magnetic fields have been inferred for a small number of giant planets from periodic stellar activity, or from transit light curves with evidence for bow shocks. Magnetic fields provide a window into interior processes such as convection, and likely regulate atmospheric escape. How escape scales with planetary and stellar properties is still not well-understood, providing an opportunity for exoplanet/solar system synergies.

Recent discoveries of nearby terrestrial planets, including HZ worlds orbiting late-type M dwarfs, have provided some of the first terrestrial targets for characterization. However, initial characterization attempts with HST, Spitzer, and ground-based telescopes have been able to provide only atmospheric constraints via nondetections of atmospheric features. Spitzer phase curves of a hot terrestrial planet that receives 70 times Earth's insolation indicates little or no atmosphere. HST and Spitzer observations of a handful of hot and HZ terrestrials, when combined with laboratory data and theory, make cloudless and cloudy hydrogen-dominated atmospheres less likely than denser ones.

Despite these early successes, there are many opportunities for improved atmospheric characterization. Today, chemical abundances are typically measured with a precision of only an order of magnitude, a sign of the still limited data quality of the challenging transit measurements, which preclude a detailed understanding of planetary formation and evolution. Atmospheric hazes and clouds in many

exoplanets further obscure the gaseous absorbers, and narrow wavelength ranges and current approximate cloud models limit our ability to account for the effects of these atmospheric aerosols. Our reduced insight into some solar system planets (particularly Venus, Uranus, and Neptune) in turn limits our understanding of the dynamics, composition, and evolution of atmospheres, indicating the need for further study of these worlds. Transmission spectroscopy can never be sensitive to the planetary surface, supporting the need for future direct spectroscopy of potentially habitable worlds. Last, atmospheric characterization has focused almost exclusively on shorter period, larger planets, and we cannot yet systematically connect atmospheric composition to the density/bulk compositional properties of longer period planets of all sizes, which often have less well characterized masses and radii.

Astrophysics Assets and Solar System Science

The planetary science community has made valuable use of astrophysics assets such as HST, Spitzer, and Kepler to explore solar system targets, which in return advance exoplanet science and astrobiology. Planetary scientists have measured the composition and orbital dynamics of small bodies to better understand solar system formation; observed diverse planetary atmospheres to assess how planetary processes are affected by composition and incident solar radiation; probed the interiors of volatile-rich bodies and identified new potentially habitable environments through the study of plumes on Europa and Enceladus; and observed the effects of extreme tidal heating on Io’s interior composition and volcanic activity. This coordination has led to discoveries that benefit both science communities.

The Dawn of Exoplanet Astrobiology: The Search for Habitable Environments and Life

In the past 10 years, exoplanet astrobiology has transformed from a field driven by promising statistical predictions to one with targets accessible to near-term observation. Significant advances have been made in our understanding of how to identify potentially habitable worlds, and how to best search for signs of life in their environments. Theory and observations now suggest that there are many evolving interactions between a planet, star, and planetary system that affect the likelihood that a planet can support a surface ocean—and that a comprehensive, systems-level approach to habitability assessment is now needed. These studies have identified systems-level challenges to habitability for M dwarf HZ planets that are less likely to be experienced by planets orbiting in the HZ of more Sun-like stars, including radiation and stellar-wind-driven atmosphere and ocean loss, and gravitational interactions that modify orbits, rotation rate, and climate.

Within the solar system, observations of Mars, Europa, Enceladus, and Titan have revealed subsurface environments that potentially harbor liquid water, and greatly expanded the ocean worlds in our solar system. Comparison of the gas giant satellites provided a systems-level view of how planetary size, formation, and tidal interactions work together to impact differentiation, ocean depths, pressures, and surface activity. These efforts forged links with the oceanography community in understanding water-rock reactions, hydrothermal vents, ocean pH, circulation, and ice/ocean interactions. Observations and missions to small bodies in the solar system illuminated processes of volatile evolution and delivery to forming planets, while exoplanet science revealed planetary system architecture influences on small body inventories and organic delivery in debris and protoplanetary disks. Venus provided context for *loss* of habitability, with relevance for Venus-analog extrasolar planets, and studies of stellar wind/planetary atmosphere interactions at Mars discovered and informed planetary atmospheric loss processes.

The astrobiological foundation needed to guide the search for signs of life’s impact on a planet’s surface and atmosphere, so-called biosignatures, has also advanced considerably. Improved understanding of the co-evolution of life with Earth environments over the past 4 Gy has highlighted how life has modified Earth’s atmosphere, surface, oceans, and interior. Life’s global impacts on a planet’s atmosphere, surface, and temporal behavior may therefore manifest as potentially detectable exoplanet

biosignatures, or technosignatures—if that life is technologically capable. Key frontiers in biosignature science now focus on the identification of novel biosignatures beyond the canonical O_2/O_3 and CH_4, especially those that are agnostic to life's molecular makeup or metabolism; understanding nonlife planetary processes that may mimic, destroy or alter potential biosignatures; and taking the first steps toward developing a comprehensive statistical framework for biosignature assessment that uses critical observables of the star, planet, and planetary system to determine the probability, and increase our confidence, that a potential biosignature is owing to life.

QUESTIONS AND DISCOVERY AREA

We now stand at a pivotal moment in our exploration of the universe, where the answers to several fundamental questions about humanity's cosmic context are within our grasp.[6] With the rapid increase of known exoplanets and the possibility of comprehensively understanding many nearby planetary systems, we can now determine if the solar system is common, or a cosmic rarity. We can understand how exoplanets form, interact, and evolve within their planetary systems, and work to understand how these interactions and processes might enable habitability on terrestrial worlds like our own. With the past decade of scientific and technical advances behind us, we now have the foundation to begin the search for habitable planets and life beyond the solar system in earnest—to address a question that humankind has been asking itself for millennia: Are we alone in the universe? To drive the strategy to explore strange new worlds, the panel has identified four questions that lead to a discovery area centered on the search for life in the universe. Each question is described in the sections that follow, and more detailed information on the capabilities required to address each question are provided in Table E.1 at the end of this appendix.

E-Q1. WHAT IS THE RANGE OF PLANETARY SYSTEM ARCHITECTURES AND IS THE CONFIGURATION OF THE SOLAR SYSTEM COMMON?

Until recently, our solar system seemed to be an orderly blueprint of a typical planetary system, with terrestrial planets interior to the snowline, gas and ice giants exterior to the snowline, and a remnant belt of unfinished planet formation at its edge. The discovery of evaporating hot Jupiters, sub-Neptunes, and eccentric gas giant exoplanets upended that notion. It is now an open question as to whether the solar system's architecture is common, or a rare outcome of chaotic dynamical evolution. In the coming decades, the study of exoplanets will expand from planets as individual objects, to interacting objects within planetary systems, presenting a new opportunity to understand our place in the universe and the histories of nearby planetary systems (compare the sections F-Q4 and Discovery Area in Appendix F).

E-Q1a. What Are the Demographics of Planets Beyond the Reach of Current Surveys?

Exoplanet demographics and occurrence rates form the foundation of our statistical understanding of exoplanetary systems, but we have not yet completed the planet census. Although the number of known exoplanets (~4000) has increased by an order of magnitude since the 2010 Decadal Survey, very few of these are analogous to solar system planets. The Kepler mission discovered thousands of short-period exoplanets, and at longer periods radial velocity (RV) surveys have primarily been limited to the most massive gas giants. The Nancy Grace Roman Space Telescope microlensing survey is poised to

---

[6] In addition to community inputs in the form of white papers and presentations, the congressionally mandated reports by the National Academies of Sciences, Engineering, and Medicine, *Exoplanet Science Strategy* and *An Astrobiology Strategy for the Search for Life in the Universe*, were considered as inputs to the panel.

greatly expand our knowledge to longer orbital periods and lower planet masses across a wide range of stellar spectral types, filling key gaps in the census, and providing a statistical anchor for planet formation and evolution models. Much like the Kepler data set revealed a gap in planet radii indicative of atmospheric evaporation, these extended demographics should give insight into physical processes governing planetary systems—for example, by detecting an enhanced density of planets near the snowlines of systems. Although microlensing surveys generally detect only a single planet in a system, these demographics will enable investigations into the architectures of planetary systems by comparison with population synthesis models. However, population-wide statistical comparisons alone are expected to leave many degeneracies in our understanding of individual systems.

E-Q1b. What Are the Typical Architectures of Planetary Systems?

By studying individual planetary systems in detail, we can understand correlations between planet populations indicative of the dynamical histories of systems, which are not expressed in statistical demographics. However, current understanding of multiplanet systems is limited to the very innermost regions, and biased toward highly coplanar systems. To put the solar system in context, we need to study regions of multiplanet systems at stellar irradiation levels comparable to those from Mercury to Neptune in our system. In the near term, our knowledge of individual planetary systems will have to be pieced together from multiple techniques; the overlap of planetary systems detectable with TESS transits and transit-timing variations (TTVs), along with RV surveys, Gaia, ALMA, and Roman Space Telescope Coronagraphic Instrument (CGI) will likely enable a clearer picture for a small number of predominantly edge-on systems. Future ground-based extremely large telescopes (ELTs) may directly image dozens of additional planets. A larger sample of well-studied nearby planetary systems would require space-based direct imaging. Such a sample would inform the range of outcomes from planet formation, constrain planet formation and evolution models, enable studies of how the system architecture may relate to past stochastic events and volatile delivery, and provide solar system context.

E-Q1c. How Common Is Planetary Migration, How Does It Affect the Rest of the Planetary System, and What Are the Observable Signatures?

While there is overwhelming evidence from planetary composition and orbital parameters that planet migration occurs, we have yet to understand the details and implications of the process (see section F-Q4 in Appendix F). Does migration commonly disrupt the rest of the system, resulting in events similar to Earth's Late Heavy Bombardment? Is there evidence of migration reversal in other systems, as proposed by the Grand Tack model of the solar system? Migration may imprint itself in the atmospheres of planets. Atmospheric elemental ratios, particularly carbon to oxygen, may record the location of formation of the planet with respect to various snowlines. Migration may also impact the composition of planetesimals and dust. Simulations and observations of the solar system suggest that Jupiter's migration may have radially mixed chemically separate reservoirs of material. Spatially resolved IR spectroscopy will enable general interpretation of disk composition, including identification of water ice and silicates; JWST will probe the warm regions of disks, while longer wavelengths are needed for colder regions. To fully understand migration, the chemical and dynamical conditions of planets must be studied prior to, during, and after migration. By combining detailed mm wavelength observations of protoplanetary disks with cold planet demographics and a large number of well-studied individual planetary system architectures, we can relate atmospheric properties and locations of planets in mature systems to those in protoplanetary disks.

E-Q1d. How Does the Distribution of Dust and Small Bodies in Mature Systems Connect to the Current and Past Dynamical States Within Planetary Systems?

The dynamical history of our planetary system is imprinted on the distribution of minor bodies in the solar system. Dynamical models of our system's past are largely able to reproduce the currently known distributions under specific conditions, suggesting that our architecture is only one of many possible outcomes. The orbital distribution and total mass of debris disks can similarly probe the past dynamics of exoplanetary systems. Panchromatic imaging from the visible to mm is necessary to study both the dust and planetesimal populations, as well as break degeneracies when modeling disk composition using unresolved spectral energy distributions.

Current mm-wavelength imaging is limited to the brightest disks, typically ~1000× the density of the Kuiper Belt. Large dust grains, observable at mm wavelengths with ALMA and next-generation radio telescopes, track the distribution of their parent planetesimals. With more debris disks resolved at mm wavelengths, via improved sensitivity to fainter disks, samples of known planet host stars and disk host stars will begin to overlap, enabling studies of dynamical interactions between exoplanets and disks (compare the Discovery Area section in Appendix F).

Smaller dust grains produced by planetesimals are observable with facilities like HST space telescope imaging spectrograph (STIS), Roman CGI, and ground-based adaptive optics (AO) coronagraphy. These grains are transported inward from the cold outer regions by radiative forces, and planets can interact strongly with them to create large-scale structures. Current and near-term visible observations of cold disks will be limited to disks ~1000× as dense as the Kuiper Belt. In these dense disks, collisions limit the types of structures that planets can create to simple ring-like belts and inclination warps; near-term observations will focus on large-scale morphology, time variability, and dust composition/optical properties. Pushing to disks ~10× as dense as the Kuiper Belt with future observations could probe a new regime of debris disk physics in which collisions subside and disks become transport-dominated. In these fainter disks, planets can imprint resonant structures that constrain planet mass and orbit, reveal the presence of otherwise undetectable planets, and help complete our picture of nearby systems (see also the Discovery Area section in Appendix F).

E-Q1e. Where Are the Nearby Potentially Habitable Planets and What Are the Characteristics of Their Planetary Systems?

Occurrence rates derived from Kepler data suggest that HZ Earth-size planets are common around M stars, and are not exceedingly rare around Sun-like (FGK) stars. Indeed, a handful of potentially habitable planets amenable to atmospheric characterization have already been identified around M type stars. HST spectroscopy has ruled out the most easily characterized (cloud-free, H/He-dominated) atmospheres for some of these transiting worlds, but further characterization via transit or high-contrast reflected-light spectroscopy should be feasible with facilities like JWST and the ELTs. In the near term, transit photometry and RV surveys will find many more of these M dwarf systems, although some may be too distant for atmospheric characterization. An extended TESS mission (or the European Space Agency's [ESA] PLATO) could also identify new HZ transiting planets around more Sun-like stars. However, to complete the census of nearby systems, detection of nontransiting potentially habitable planets around Sun-like stars will require improved RV sensitivity and space-based direct imaging. Solar observations and collaboration with heliophysicists may be fruitful for understanding and modeling the effects of stellar variability on RV observations (compare the section G-Q3 in Appendix G).

In addition to finding these nearby planets, we must understand the systems within which they reside. Studying their planetary systems could reveal correlations and key processes that impact habitability. For example, the specifics of Jupiter's past migration may have largely determined the architecture of and volatile delivery in our solar system. Accordingly, looking at systems with Jupiter analogs might be a way to find systems with architectures broadly similar to that of the solar system.

E-Q2. WHAT ARE THE PROPERTIES OF INDIVIDUAL PLANETS AND WHICH PROCESSES LEAD TO PLANETARY DIVERSITY?

We know from decades of solar system exploration that planets are individually complex and collectively diverse. A complete understanding of planets requires appreciation of not only their physical properties, but also the underlying processes that shape them. Thus, the goal of exoplanet characterization is to measure their atmospheric, surface, and internal compositions, and to infer their radiative, chemical, dynamical, and magnetic field processes. Comparisons between multiple planets within single planetary systems are particularly helpful for illuminating such processes. While exoplanet science has only scratched the surface of our understanding of individual planets and their properties, the next decade will allow more detailed and meaningful characterization of a diversity of planets. Important goals for the coming decade include addressing the following questions.

E-Q2a. Which Physical Processes Govern a Planet’s Interior Structure?

Exoplanet science has been guided by the mass-radius relationship, which helps constrain planetary bulk density. While general trends are apparent, important questions remain, including why planets with similar masses have different densities and how their interior composition can be inferred. While Uranus and Neptune have notably different densities and internal heat flows despite their similar mass, the diversity of exoplanet densities in this mass range, albeit at higher stellar irradiation than Neptune and Uranus, is particularly large, suggesting an even larger diversity of interior structures and compositions. Placing solar system planets within the context of the intrinsic diversity of all exoplanets, including the sub-Neptune and super-Earth-size planets that have no solar system analogs, will illuminate how bulk planetary properties, and formation and thermal histories, affect planetary interior structure and magnetic fields. These studies will be informed by larger samples of planet radii and masses at higher precision (particularly for cooler planets than yet characterized), improved theoretical approaches, and new laboratory measurements of the equations of state and chemical properties of planetary materials at high temperatures and pressures.

E-Q2b. How Does a Planet’s Interior Structure and Composition Connect to Its Surface and Atmosphere?

The atmosphere of a planet is not necessarily a tracer of its bulk composition. A planet may have discrete compositional layers of which the atmosphere is only the outermost. Furthermore, models suggest that the observable atmospheres of transiting hot Jupiters are separated from deeper atmospheric layers by a radiative layer that may inhibit mixing, disconnecting the atmosphere from the deeper interior. Condensation, circulation patterns, and various sources of chemical disequilibrium likely also affect the composition of the remotely detectable atmosphere. For terrestrial planets, surface/atmosphere exchange mechanisms mediate atmospheric composition, and planetary magnetic fields can illuminate processes occurring deep in a planet’s interior, while providing critical insights into how the planet’s atmosphere interacts with the space environment. Meeting the goal of determining the bulk composition of a planet thus entails connecting the observable atmosphere, as sculpted by such processes, to deep atmospheric or surface processes and chemical composition. Theory and laboratory studies inform our understanding of the deep atmospheric process. Statistical surveys of atmospheric composition, such as by ESA’s ARIEL mission, and bulk planetary properties will illuminate the diversity and trends. For example, do lower mass gas giant planets exhibit consistently higher atmospheric enrichment in heavy elements than do higher mass planets? High-resolution spectroscopy will certainly be a major capability of the proposed ELTs, which would provide particularly robust measurements of molecular species and thermal structure in both short-period and directly imaged planets. These data will be complemented by transmission and

emission spectra of transiting planets spanning nearly the entire range of exoplanet masses, sizes, and temperatures, especially with the expanded spectral range and enhanced sensitivity of JWST. More precise high-frequency radio observations can be used to increase the sample of planets known to have magnetic fields, and lower-frequency observations can detect the weaker magnetic fields that are more likely to be present in ice giant and smaller gas giants. With larger observational samples, more sensitive observing facilities, and better theoretical and laboratory insights, we can hope to meaningfully connect observed planetary characteristics to formation and evolutionary history. A better understanding of structure and composition of solar system planets—especially Venus and the ice giants—will inform these exoplanet studies. Including exoplanet scientists as team members and science investigators on future missions to Venus, Uranus, and Neptune would advance efforts to understand analog exoplanets.

E-Q2c. What Fundamental Planetary Parameters and Processes Determine the Complexity of Planetary Atmospheres?

The current state of a planet's atmosphere depends not only on its formation environment and co-evolution with its planetary interior through such processes as interior outgassing and magnetic field generation but also on ongoing processes such as photochemistry, cloud formation, and atmospheric dynamics. A clearer understanding of a planet's evolutionary pathway on all time scales demands substantial efforts in both observations and theoretical modeling. Phase-resolved observations from facilities such as JWST, and high-dispersion spectroscopy conducted from the ground, ideally in tandem with climate, photochemistry, and 3D atmospheric dynamical models, provides estimates of the atmospheric composition and dynamics, and insights into planetary rotation state. The microphysical and dynamical processes that govern the morphology and transport of clouds have yet to be untangled for giant exoplanets, and detailed models of these processes will be imperative for understanding formation and transport of clouds on all types of terrestrial planets with atmospheres. Polarization observations could lend further insights. Photochemistry may produce observable features in transmission or reflected spectra, and are likely to be sensitive to stellar UV output, atmospheric composition, and dynamical processes. Panchromatic stellar characterization (see the sections G-Q1 and G-Q3 in Appendix G), laboratory experiments, and studies of solar system analogs will support studies in these areas.

E-Q2d. How Does a Planet's Interaction with Its Host Star and Planetary System Influence Its Atmospheric Properties over All Time Scales?

A planet's external environment—the star (see section G-Q4 in Appendix G) and other bodies in the system—also plays a critical role in shaping the evolution of its atmosphere. The sub-Neptune exoplanet radius gap has been interpreted as owing to strong hydrodynamic escape processes driven by the star. A growing number of exoplanets exhibit evidence for active escape processes driven by stellar photon fluxes and stellar winds, and it is anticipated that stellar activity in the form of flares and coronal mass ejections can dominate atmospheric processes for some systems. The stellar UV spectrum influences planetary atmospheric photochemistry, which can modify atmospheric composition and atmospheric loss, with subsequent impacts on planetary climate. Obtaining both transit and directly imaged observations of planets, particularly (but not exclusively) at UV wavelengths that are sensitive to upper atmospheric processes, will illuminate escape processes and time scales, and provide crucial inputs for modeling. Measurements of escape and photochemistry for solar system planets reveal the variety of atmospheric escape processes, and validate models for atmospheric escape from exoplanets. Other known "star-planet" interactions include aurora (observed on brown dwarfs but not yet exoplanets) and direct magnetic connection similar to that between Jupiter and its satellites (reported for multiple systems). Beyond stellar output, planetary impactors, perhaps mediated by other planets in the system, deliver and remove

volatiles; and star-planet-planetary system gravitational interactions can induce tidal heating and enhance outgassing.

E-Q2e. How Do Giant Planets Fit Within a Continuum of Our Understanding of All Substellar Objects?

The thousands of known brown dwarfs span the gulf in mass and temperature from the smallest main sequence stars to cold gas giant planets like Jupiter and Saturn. Brown dwarfs are easier to study in detail than planets, offering the opportunity to rigorously test models of thermal evolution, atmospheric dynamics, chemistry, cloud formation, and magnetic dynamos that are also applicable to extrasolar giant planets. Characterizing the differences and similarities of the two classes of objects will elucidate their formation mechanisms, informing the limits of both planet and star formation (compare section G-Q1 in Appendix G).

Young, low-mass brown dwarfs serve as particularly valuable laboratories for refining models of young directly imaged planets because they have similar gravities and temperatures. Roman CGI will obtain optical wavelength thermal emission spectra of young companion objects as a complement to JWST and ground-based, longer wavelength observations, and perhaps reflected light spectra of a few cool giants. Both types of observations will inform the properties of giant planets, helping to place them in context with low-mass brown dwarfs. Additional surveys for substellar companions to stars that probe to higher contrasts and smaller separations will find more young, planetary mass companions.

While many individual and binary brown dwarf systems are known, the relatively rare companions to main sequence stars allow comparison of the companion's composition to that of its primary star. Future observations of the masses and atmospheric and magnetic properties of brown dwarfs and giant planets will reveal how substellar companions to stars form (compare section G-Q2 in Appendix G) and evolve, and whether the processes are similar for giant planets and brown dwarfs over the observable ranges of mass and orbital separation. Progress here requires improved spectral data, including higher resolution and greater wavelength coverage in polarized and unpolarized light, astrometric and/or RV masses for many more substellar companions, and improved theoretical modeling approaches and laboratory data.

E-Q3. HOW DO HABITABLE ENVIRONMENTS ARISE AND EVOLVE WITHIN THE CONTEXT OF THEIR PLANETARY SYSTEMS?

The habitability of a planet is governed by a complex interplay of planet, star, and planetary system architecture and the mutual evolution of these components over time. Consequently, the context provided by the host star and planetary system architecture, including the distribution of small bodies and their potential for volatile delivery, is important for determining whether a habitable planet can form and maintain its habitability over time. An improved understanding of the factors and processes influencing habitability are needed to support exoplanetary exploration and target selection. To identify habitable environments and connect them to the planetary systems in which they reside, foundational research on exoplanet properties and processes through observations of planets, disks, and planetary systems and theoretical models, laboratory studies, and comparisons with solar system analogs is needed.

E-Q3a. How Are Potentially Habitable Environments Formed?

A planet acquires volatiles and organics essential for a habitable environment either during formation and migration or via subsequent impacts of volatile and organic-rich bodies. Factors that affect the type and amount of volatiles acquired include the type of star, its metallicity, the composition of the

disk that formed around the star during planet formation, and migration of planets within the system after bodies form. Answering this question requires an improved history of volatiles in our solar system, characterization of volatiles in exoplanet systems, and modeling how volatiles are acquired and lost by potentially habitable planets. Completing the volatile and organic inventory of planets within the solar system, including the dwarf planets, asteroids, Kuiper Belt objects (KPOs) and comets, and determining the dynamical interactions that formed these populations, provides the tightest constraints on the properties of the protoplanetary disk from which our system formed. These studies would also help determine how volatiles were distributed during and after planet formation and migration, and could reveal how volatiles are incorporated in forming and evolving planets. Complementary measurements of volatile content across planet forming regions in exoplanetary systems could provide context for how and when major dynamical events took place within the solar system, and reveal the photochemical processes that gave rise to known small-body compositions. These studies could constrain theoretical models of dynamical evolution, and volatile accretion and delivery, informing how architecture, composition, and timing interact to determine which planets acquire volatiles.

E-Q3b. What Processes Influence the Habitability of Environments?

Once a volatile- and organic-rich planet is formed, acquiring and maintaining an atmosphere and surface ocean relies on a suite of planetary, stellar, and planetary system properties and interactions. These include star/planet interactions that govern the loss or maintenance of a primordial planetary atmosphere (see section G-Q4 in Appendix G), and the planetary interior/atmosphere exchanges that can generate and replenish a secondary atmosphere and ocean. Taking a systems science approach to habitability will strengthen our understanding of these different processes for planets of different compositions. Within the solar system, processes like tidal heating, asteroid bombardment, and the loss and evolution of planetary atmospheres through escape and photochemistry can be studied to better understand their impact on solar system terrestrial planets and ocean worlds. Additionally, Earth's interior, surface, and atmosphere have evolved significantly over its history owing to a wide range of geological, photochemical, and biological processes, providing a range of different habitable environments over time. Combination of measurements and theory of the nature and processes that drove Earth's early habitability, and the loss of habitability on Venus and Mars can inform our understanding of exoplanet habitability. The subsurface ocean worlds of Europa, Enceladus, and Titan likely also harbor habitable environments that are governed by processes, such as tidal heating, that may also be relevant to habitable exoplanets. To support the insights provided by the systems science approach and solar system analogs, strongly interdisciplinary work is needed between planetary science, astronomy, Earth science, and heliophysics/stellar astronomy, including laboratory and theory. In the longer term, studying Earth-size planets near the HZ of other stars (see section E-Q3d, below) will provide observational insights into the characteristics and processes of a broader range of habitable planets.

E-Q3c. What Is the Range of Potentially Habitable Environments Around Different Types of Stars?

Earth, orbiting a G dwarf, is the only habitable planet known to support life. However, exoplanets coevolve with their host stars, just as Earth coevolved with the Sun. The host star impacts planetary atmospheric loss, composition, and climate, and the host star's spectrum (including X-ray/EUV flux), activity, and long-term luminosity evolution are critically important for understanding the dynamic habitability of exoplanets. Exoplanet surveys have shown that terrestrial planets can exist around a range of stellar types, but observations have yet to confirm if habitable environments can exist around all types of stars. Although M dwarf planets will be the first accessible to near-term observation, they are far more likely than Sun-like stars (FGK dwarfs) to drive planetary atmosphere and ocean loss. The close-in HZ makes M dwarf HZ planets potentially more vulnerable to atmospheric loss, coronal mass ejection events,

tidal heating, and orbital evolution. To better understand the distribution of habitable environments in the local solar neighborhood, it is important to understand how the star's properties and evolution influence the evolution and habitability of terrestrial planets (compare sections G-Q2, G-Q3, and G-Q4 in Appendix G). To improve our understanding of these processes, stellar energetic output for a range of spectral types and over multiple temporal scales is needed, combined with theoretical models of both magnetized and unmagnetized planets to understand the impact of the host star on atmospheres and oceans for a large sample of systems and spectral types.

E-Q3d. What Are the Key Observable Characteristics of Habitable Planets?

Modern Earth provides the only observable example of a habitable surface environment. To expand our understanding of the observational discriminants for habitable environments, we need to study Earth's environments through time, relevant solar system environments, and both model the observable characteristics of, and ultimately observe, potentially habitable environments under the influence of different types of stars. For exoplanets, initial observational assessment of habitability requires determining the presence and nature of an atmosphere, and searching for atmospheric or surface signs of the presence of an ocean. In the near-term, observations of M dwarf HZ planets with JWST and ELTs could identify the presence of atmospheres and detect key molecules that could make habitability more or less likely. In the longer term, direct imaging mapping of phase-dependent ocean glint could directly show the presence of an ocean, although the likelihood of habitability could also be inferred from observations and theory constraining the surface conditions. Strongly interdisciplinary efforts, combining observations, laboratory, and theoretical studies, are needed to study and identify signs of habitability prior to future observations.

E-Q4. HOW CAN SIGNS OF LIFE BE IDENTIFIED AND INTERPRETED IN THE CONTEXT OF THEIR PLANETARY ENVIRONMENTS?

Over the next 10 years, JWST and upcoming ground-based telescopes will have the opportunity to conduct the first searches for signs of life on terrestrial planets orbiting a handful of nearby M dwarf stars. However, detecting potential signs of life with upcoming technology is only one component of our search for life. To support these efforts, strongly interdisciplinary science is also required to identify which biosignatures to look for, and to understand how to assess whether a potential biosignature is more or less likely to be owing to life, given the context of the planetary environment.

E-Q4a. What Biosignatures Should We Look For?

Astrobiologists in solar system and exoplanetary science have worked together to identify a short list of proposed atmospheric, surface, and temporal exoplanet biosignatures—based largely on our modern Earth and the past environments and dominant metabolisms of early Earth. An ideal biosignature must satisfy three major criteria—it must be reliably produced by life, must survive or be preserved in its environment, and must be detectable with anticipated technology. Under these criteria, the broad global impacts of the harnessing of abundant sunlight by oxygenic photosynthesis, remain a key set of biosignatures. However, to increase the probability of finding and recognizing life elsewhere, we need to continue identifying alternative biological pathways that could produce detectable biosignatures, and explore the potentially detectable impacts on a planetary environment by "life as we don't know it." The latter goal would be met by developing the new frontier of "agnostic biosignatures" that are not associated with a specific metabolism, but may take the form of unanticipated complexity in a planetary environment, as revealed by atmospheric chemical networks or disequilibria. To propose potential

biosignatures that are more likely to be detectable, we also need to observe and model how processes in the interiors, surfaces, and atmospheres of planets can work to enhance or destroy a biosignature, or how abiotic processes might mimic a biosignature and complicate its interpretation. Coordinated work by heliophysicists/stellar astronomers (compare sections G-Q3 and G-Q4 in Appendix G), biologists, and planetary and Earth scientists is needed to combine observations of Earth and solar system planets, laboratory work, and theory to identify novel atmospheric, surface, temporal, and agnostic biosignatures accessible by upcoming missions.

E-Q4b. How Will We Interpret the Biosignatures That We See?

Recent advances in astrobiology research have shown that it is likely that all biosignatures, including abundant O_2, O_3, and CH_4, will need to be interpreted in the context of their planetary environment to rule out false positives—planetary processes that could mimic the biosignature. Modeling of star-planet interactions suggests that O_2 may have abiotic production mechanisms, including photochemistry and ocean loss, which are especially likely for planets orbiting M dwarfs. Consequently, any potential biosignature (or technosignature) will need additional assessment to determine whether it is more likely to have a biological origin, by using environmental context to rule out false positives, and to search for secondary confirmation of the biosignature hypothesis. Over the next decade, we will need to develop a comprehensive framework for probabilistic biosignature assessment to determine whether the observed phenomenon is more or less likely to be owing to life. For each biosignature considered, such a framework would need to consider the context of the stellar and planetary environment, and include an understanding of the false negatives, false positives, and observational discriminants. To support this framework, observations of a host star's UV spectrum and activity (an overlap with section G-Q4 in Appendix G), and a wide range of planetary types, from gas giants to uninhabitable terrestrials, will need to be combined with theoretical modeling and insights from solar system planets to improve our understanding of the physical and chemical processes that modify planetary environments. In particular, an empirical census of atmospheres on terrestrial worlds under a wide range of conditions, both in and outside the HZ, will be needed to validate or adjust current ideas about atmospheric signatures produced through abiotic and biotic processes. Additionally, we can use solar system spacecraft data to identify if the surfaces or atmospheres of solar system bodies show evidence for chemistry and organic products and processes, and whether these suggest biogenic or prebiotic potential, or constitute false positives.

E-Q4c. Do Any Nearby Planets Exhibit Biosignatures?

The next decade will present several opportunities to characterize terrestrial exoplanets and undertake the very first search for biosignatures on a handful of planets orbiting nearby M dwarfs. Owing to their host stars' super-luminous pre-main sequence phase, activity, and the proximity of the HZ to the star, M dwarf planets likely undergo a very different evolutionary history—which may include atmosphere and ocean loss—than planets orbiting more Sun-like stars, and may allow us to expand our understanding of biospheres for different stellar hosts. JWST can likely detect CO_2 and CH_4 on transiting HZ planets orbiting a few late-type M dwarfs, and so will search for biologically induced disequilibrium conditions that may have been prevalent on the early Earth. TRAPPIST-1 d, e, f, and g are likely to be the most promising targets for such searches. ELTs will complement *JWST's* initial assay by accessing a larger sample (~10) of nearby earlier-type M dwarf planets using adaptive optics techniques and/or high-resolution spectroscopy to search for O_2, which is unlikely to be detected with JWST. The detection of O_3 in the atmospheres of HZ M dwarf planets is unlikely with either JWST or the ELTs, but more precise transmission spectroscopy (~5 ppm sensitivity) could detect it in the MIR. However, transmission spectroscopy cannot probe the near-surface atmosphere and planetary surface and may miss more UV-

labile biosignature molecules, such as volatile organic compounds, that are better preserved in the near-surface environment.

DISCOVERY AREA: THE SEARCH FOR LIFE ON EXOPLANETS

Is there life elsewhere in the universe? This profound question has echoed down through the millennia, and the answer is now within our scientific and technical grasp. Ground-based surveys have transformed the search for life from a philosophical question to a near-term scientific observable by providing a handful of high-priority M dwarf terrestrial planets amenable to spectroscopic atmospheric characterization with JWST and the ELTs. However, these host stars may present challenging environments for life, be more likely to generate false positives for biosignatures, and interpretation of their planets may need substantial extrapolation of solar system-informed knowledge of habitability. A robust search for life therefore requires surveying the HZs of Sun-like stars, where we know that life can arise. This is only possible with a large high-contrast direct-imaging space telescope.

Directly imaging and obtaining spectra of objects 10 billion times fainter than their host stars is a remarkable challenge. However, the past decade has seen significant reductions in the two largest sources of astrophysical uncertainty for these observations. First, the Kepler mission has shown that roughly Earth-size planets in the HZ of Sun-like stars are not exceedingly rare, with an estimated occurrence rate of ~0.1. Second, the LBTI HOSTS debris disk survey indicates that warm exozodiacal dust is not prohibitively bright, being typically just a few times that of the zodiacal cloud; observations with the Roman CGI may also improve exozodiacal dust constraints. Terrestrial exoplanets now appear to be common enough to detect in substantial numbers with sufficient resources.

Maximizing Our Chances of Finding, Recognizing, and Quantifying Life

There are multiple ways to maximize our chances of finding life, including searching the HZ of a wide variety of stellar spectral types, being sensitive to biosignatures over a wide range of evolutionary history, probing the deepest levels of a planet's atmosphere where a larger range of biosignatures may persist, and increasing the chances of observing molecules that reveal biosignatures and environmental context. The following suite of parameters and capabilities will maximize the probability for life detection along these different axes.

Large Sample Size

The sample size of HZ terrestrial planets increases our chances of observing life, and improves our ability to quantitatively answer whether there is life elsewhere in the universe. A large sample size provides robustness against remotely detectable life being an unlikely outcome—for example, if remotely detectable life arises on 10 percent of HZ terrestrials, 30 such planets must be surveyed to detect such life with 95 percent confidence. For a null result, a larger sample size places a stricter constraint on how often remotely detectable life *could* arise—for example, if 30 HZ terrestrials are surveyed and *none* exhibit signs of life, we can conclude with 95 percent confidence that remotely detectable life arises on fewer than 10 percent of HZ terrestrials. Broadly speaking, dozens of habitable planets are required to provide an informative null result. Both the odds of "yes" and the scientific impact of "no" are increased.

Diverse Stellar Sample

Although JWST and ground-based telescopes will soon allow a tantalizing first attempt at the search for life on ~10 planets around M dwarfs, these stars present many challenges to habitability and biosignature interpretation. Earth and our G dwarf Sun are the only known planet-star combination to host life. A more definitive and informed answer to the question of whether we are alone will require searching stars spanning a broad range of spectral types, including the more Sun-like FGK stars. By using

a larger sample size that includes a range of FGKM stars, we improve our chances of finding inhabited planets, and understanding how the stellar environment impacts them.

Direct Imaging

While transmission observations will likely work well for M dwarf planets, direct imaging is needed to study the atmospheres of planets orbiting Sun-like FGK stars. Importantly, direct imaging probes the lower atmosphere near the surface, the most useful region for biosignature and ocean detection, the latter phase-dependent observations of ocean glint. In the longer-term, mid-IR interferometric imaging would provide a valuable complement to visible-NIR imaging observations.

UV Capability

O_2 reveals the presence of a photosynthetic biosphere on our planet, but it was likely only present at directly detectable levels for the past 1–2 Gyr of Earth's history. Prior to that, the presence of low levels of O_2 could have been inferred from the strong UV absorption feature of O_3. UV observations therefore critically enhance sensitivity to signs of photosynthesis over a larger range of a planet's lifetime.

Multiple Spectroscopic Bands and Species

We can increase our chances of detecting life by having the capability to detect multiple potential biosignatures (e.g., O_2, O_3, CH_4), including those produced by a range of metabolisms other than oxygenic photosynthesis. Detection of a putative biosignature gas is more robust if multiple spectral bands are detected. Similarly, interpretation of biosignature gases to assess false positive scenarios requires context on the planetary environment, including atmospheric and surface characterization. A broad wavelength range enables the detection of multiple key species, and potentially multiple bands of those species, as well as providing better constraints on any atmospheric aerosols, all of which increase the robustness of biosignature detection and interpretation.

The capabilities needed to study the environments and possible biospheres of habitable planets orbiting more Sun-like stars are not met by any existing or currently selected facilities, but developing the scientific community and technological capabilities required to do so would enable huge advances in multiple aspects of exoplanet science and astrophysics. This Discovery Area would benefit significantly from collaboration across disciplinary boundaries, and ongoing support for enabling observations, theory and laboratory work. The search for life on exoplanets will provide a bold and unifying vision for exoplanets, astrobiology and solar system science.

BOX E.1 Summary of Science Questions

E-Q1: What is the range of planetary system architectures and is the configuration of the solar system common?	E-Q1a: What are the demographics of planets beyond the reach of current surveys? E-Q1b: What are the typical architectures of planetary systems? E-Q1c: How common is planetary migration, how does it affect the rest of the planetary system, and what are the observable signatures? E-Q1d: How does the distribution of dust and small bodies in mature systems connect to the current and past dynamical states within planetary systems? E-Q1e: Where are the nearby potentially habitable planets and what are the characteristics of their planetary systems?
E-Q2: What are the properties of individual planets, and which processes lead to planetary diversity?	E-Q2a: Which physical processes govern a planet's interior structure? E-Q2b: How does a planet's interior structure and composition connect to its surface and atmosphere? E-Q2c: What fundamental planetary parameters and processes determine the complexity of planetary atmospheres? E-Q2d: How does a planet's interaction with its host star and planetary system influence its atmospheric properties over all time scales? E-Q2e: How do giant planets fit within a continuum of our understanding of all substellar objects?
E-Q3: How do habitable environments arise and evolve within the context of their planetary systems?	E-Q3a: How are potentially habitable environments formed? E-Q3b: What processes influence the habitability of environments? E-Q3c: What is the range of potentially habitable environments around different types of stars? E-Q3d: What are the key observable characteristics of habitable planets?
E-Q4: How can signs of life be identified and interpreted in the context of their planetary environments?	E-Q4a: What biosignatures should we look for? E-Q4b: How will we interpret the biosignatures that we see? E-Q4c: Do any nearby planets exhibit biosignatures?
Discovery Area	The search for life on exoplanets.

TABLE E.1 Capabilities

Capability	**Science Enabled**	**Current/Expected Facilities**	**Future Needs**
Large-aperture, space-based direct imaging	E-Q2a, E-Q2b, E-Q2c, E-Q2e, E-Q3b, E-Q3d, E-DA		Large-aperture space-based UV-NIR imaging/spectroscopy (0.3–1.8 microns, contrast ~ 1e-10, IWA < ~60 mas, OWA ~ 500 mas, R ~ 150 spectroscopy[a] for dozens of potential Earth analogs in the HZs of Sun-like stars
Radial velocity observations	E-Q1, E-Q2a, E-Q2b, E-Q2e, E-Q3d, E-DA	Ground-based PRV facilities (e.g., NEID): cold gas giants in TESS systems, masses of planets with known radii including potentially habitable planets orbiting M dwarfs; detection of long-period planets (P = 1–100 yr); using solar observations to understand effect of stellar variability on exoplanet observations	EPRV (ground and/or space; 10 cm/s semi-amplitude sensitivity for P = 50–400 days): masses and orbits of habitable planets orbiting FGKM stars
High-contrast imaging	E-Q1, E-Q2a, E-Q2c, E-Q3b, E-Q3c, E-Q3d, E-Q4c, E-DA	GPI, SPHERE, GRAVITY, Roman CGI: few gas giants, dozens of exozodiacal/debris disks ELTs: detection of habitable zone Earth-size planets around M stars (0.5–1.8 microns, contrast ~ 1e-8, IWA < ~30 mas, OWA ~ 200 mas, ~ dozen targets)	Space-based high-contrast imaging: full planetary systems including faint debris disks (visible wavelengths, contrast ~ 1e-10, IWA < ~60 mas, OWA > ~1", spatial resolution < ~0.01 mas, ~100 planetary systems)
High-contrast spectroscopy	E-Q1e, E-Q2, E-Q3b, E-Q3c, E-Q3d, E-Q4c, E-DA	ELTs: characterization of habitable zone Earth-size planets around M stars (0.5–1.8 microns, contrast ~ 1e-8, IWA < ~30 mas, OWA ~ 200 mas, R > 1e5, dozens of targets) to search for biosignatures	Space-based UV-NIR spectroscopy: characterization of HZ Earth-size planets around FGK stars (0.3–1.8 microns, contrast ~ 1e-10, IWA < ~60 mas, OWA ~ 500 mas, ~100s of stars, R ~ 150 spectroscopy for potentially dozens of Earth analogs)
Astrometry	E-Q1, E-Q2a, E-Q2b, E-Q2c, E-Q2e, E-Q3d, E-DA	Gaia, Roman WFI supplement: population studies overlapping with Kepler, cold gas giants in TESS and nearby systems	Near-IR astrometry to measure substellar object masses/orbits; masses and orbits of temperate planets orbiting FGKM stars
Polarization	E-Q1d, E-Q1e, E-Q2c, E-Q2e, E-Q3c, E-Q3d, E-Q4, E-DA	Roman: polarization of disks Ground-based instruments, including on ELTs: polarization signatures of disks and giant planets	Direct imaging to probe polarized ocean glint on terrestrial planets
Microlensing	E-Q1a	Roman population studies	

Transit observations	E-Q1b, E-Q1c, E-Q1e, E-Q2a	TESS: discover and measure radii of inner planets, evaporated cores, and migrated planets orbiting bright stars and potentially habitable planets orbiting KM stars PLATO: find planets orbiting bright stars; precisely determine planet and star properties; asteroseismology	Large collecting area: detection of extremely small (terrestrial, <1.6 Earth radius) transiting objects, exomoons, and planetary ring systems
Transit spectroscopy—O/NIR/MIR	E-Q1c, E-Q1e, E-Q2, E-Q3d, E-Q4c	JWST: characterize atmospheres of a few potentially habitable planets orbiting M dwarfs JWST, ARIEL: NIR-MIR spectra of a few mature Jovians and dozens of Jupiter to >2 R_Earth close-in planets (TESS planets) HST: atmospheric composition of warm/hot gas giants	Large collecting area ground or space telescopes: extremely high SNR transit observations to study atmospheric dynamics and variability; wavelength coverage out to 20 micron; spectrophotometric stability in transit of <10 ppm below 10.5 micron, and <25 ppm above 10.5 micron
Stellar characterization	E-Q2d, E-Q3b, E-Q3c, E-Q3d	HST, Chandra, XMM: EUV/NUV/O/IR spectra of known planet host star; important to have multiple lines and multiple bandpasses (e.g., Lyman alpha and Mg II in UV, and X-ray for CMEs) UV: atmospheric escape	X ray: More sensitive observations (~50× greater than Chandra) with R > 5000 and high spatial resolution UV: unresolved UV monitoring of nearby FGKM stars and time-resolved stellar spectra IR: high spectral resolution (R > 20,000) Radio observations > 10 GHz and high-resolution Lyman alpha (> ~ 30,000) for photoevaporation and inferring stellar mass loss rates
UV observations of planets and host stars	E-Q2a, E-Q2c, E-Q2d, E-Q3c	HST limited UV transit capability	UV space telescope: R > 1000 spectroscopy; monitor atmospheric escape; high-contrast imaging of planets to detect UV absorbers; time-resolved UV stellar flux
High-resolution O/IR spectroscopy	E-Q2a, E-Q2.c, E-Q2d, E-Q2e, E-Q4c	R > ~1e5 O/IR spectroscopy (8–10 m telescopes): giant planet characterization (~dozens) HST, Keck: UV-VIS-NIR transit spectroscopy for atmospheric characterization/escape ELTs: VIS-NIR spectroscopy (transmission and reflected light) to detect O_2 in M dwarf terrestrial exoplanet atmospheres (0.5–1.8 microns, contrast ~ 1e-8, R > 1e5, dozens of targets)	High-contrast vis-NIR reflected light spectroscopy of mature planets Coupling of AO and high-dispersion spectrographs on ELTs

Emission photometry and spectroscopy	E-Q2	Large ground-based telescopes; JWST; ARIEL Eclipse, phase curve spectra (JWST): near- to mid-IR spectra of dozens of >2 R_Earth transiting planets.	R ~ 100 MIR eclipse spectroscopy of temperate/cooler Neptune and larger planets
Long-baseline mm interferometry	E-Q1a, E-Q1c, E-Q1d, E-Q2a	ALMA: ~dozen planetesimal belts around known massive disks to measure belt locations/geometries/masses, many detailed characterizations of bright protoplanetary disks	High-sensitivity mm interferometry: ~10–100× improved sensitivity to image dozens of SS-like planetesimal belts (1 mm wavelength, resolution < ~1", sensitivity ~ 0.1 microJy/beam)
Long-baseline, long-wavelength interferometry	E-Q2d	JVLA	Low-frequency radio arrays with several mJy sensitivity from ~50 MHz: detecting radio emission from magnetic fields of exoplanets
Mid-IR direct imaging	E-Q2a, E-Q2b, E-Q2c, E-Q2e, E-DA	ELTs: 10 micron high-contrast imaging from ground for a few stars to measure radii and temperatures of few planets	Mid-IR interferometry: measure temperature, radii, and atmospheric features of planets around FGKM stars, including Earth-size HZ planets (5–18 micron)
Solar system small-body characterization	E-Q1b, E-Q1c, E-Q1d	HST, ground-based telescopes	Large time allocations and/or improved detection algorithms; continued detection and spectroscopic characterization of small bodies in UV/IR; rotation rates and orbital characteristics; small KBO binary fraction
Characterization of solar system planets	E-Q2, E-Q3b, E-Q3d, E-Q4b	Venus atmospheric composition Atmospheric escape Mars	Ice giants: atmospheric and interior structure and composition Venus: atmosphere entry probes
Habitability relevant solar system environments	E-Q3b, E-Q4a	Dragonfly: Titan; Europa Clipper: Europa JUICE: Galilean moons Mars2020: Mars Ground-based observations: Venus	Venus: atmospheric chemistry assays Enceladus: future missions Earth: detectable characteristics of Earth environments through time
Interdisciplinary theory, laboratory, field	E-Q3, E-Q4a, E-Q4b, E-DA	*Identification of novel biosignatures, comprehensive multifactorial framework for habitability assessment, identification of biosignature false positives and negatives and their observational discriminants. Probabilistic framework for biosignatures assessment. Always needed.*	
Laboratory studies	E-Q2, E-Q4b, E-DA	*Planetary interiors: volatile solubilities in planetary materials, equations of state, high-pressure melting curves, viscosities, thermal conductivities.* *Planetary atmospheres: composition, UV-MIR opacities, and other properties of gases/aerosols/particles for atmospheres and disks.*[b] *Photochemical and ion rate reactions. Always needed.*	

Theory	E-Q1c, E-Q2, E-Q3b, E-Q4a, E-Q4b, E-DA	*Simulations of orbital evolution, migration, and interactions of planets with small bodies. Modeling of planetary interiors and interior-surface-atmosphere-magnetosphere exchange. 1D to 3D atmospheric models (including star-planet interactions) evaluating photochemistry and haze formation, clouds, dynamics, climate, and escape; modification of planetary atmospheres over daily, seasonal, stellar cycle, and evolutionary time scales. Computational molecular opacities and line profiles. Always needed.*
Cross-division data analysis programs	E-Q1b, E-Q1c, E-Q1d, E-Q2d, E-Q3a, E-Q3b, E-Q3c, E-DA	*Programs that enable and expand opportunities for synergistic exoplanet/solar system/Earth science/heliophysics research and interactions. Always needed.*
Cross-division mission participating scientist programs	E-Q2, E-Q3b, E-Q3c, E-DA	*Opportunities for participation by exoplanet scientists in heliophysics, Earth science and solar system exploration missions, and the participation by planetary scientists, Earth scientists, and heliophysicists in exoplanet relevant astrophysics missions.* *Not currently provided by NASA R&A program structure. Always needed.*

[a] T.D. Brandt and D.S. Spiegel, 2014, Prospects for detecting oxygen, water, and chlorophyll on an exo-Earth, *Proceedings of the National Academy of Sciences* 111(37): 13278–13283; Y.K. Feng, T.D. Robinson, J.J. Fortney, R.E. Lupu, M.S. Marley, N.K. Lewis, B. McIntosh, and M.R. Line, 2018, Characterizing Earth analogs in reflected light: atmospheric retrieval studies for future space telescopes, *Astronomical Journal* 155(5):200.
[b] For example, J.J. Fortney, T.D. Robinson, S. Domagal-Goldman, D.S. Amundsen, M. Brogi, M. Claire, M.S. Marley, et al., 2016, The need for laboratory work to aid in the understanding of exoplanetary atmospheres, arXiv preprint arXiv:1602.06305.
NOTE: IWA/OWA: inner and outer working angles for optimum starlight suppression in direct imaging systems; R: spectral resolution; P: planetary orbital period; CGI: coronagraphic instrument; E/PRV: extreme/precision radial velocity; NEID: NN-EXPLORE exoplanet investigations with Doppler spectroscopy; GPI: giant planet imager; SPHERE: Spectro-Polarimetric High-Contrast Exoplanet Research; Roman WFI: Roman Wide-Field Imager; CME: coronal mass ejection.

F

Report of the Panel on the Interstellar Medium and Star and Planet Formation

INTRODUCTION

Stars and planets form from gas and dust initially present in the interstellar medium (ISM). The ISM is complex, highly structured, and dynamic. It consists of gas at temperatures ranging from ~10^7 K or more down to 10 K or lower, with densities ranging over many orders of magnitude, threaded by magnetic fields. Stellar energy input in the form of radiation, winds, and explosions shape the ISM, heating, ionizing, and dissociating atomic and molecular gas, driving the cycling of material through the different phases, dispersing and accumulating dense gas, and even expelling gas into the circumgalactic medium (CGM). Stellar winds and explosions enrich the ISM in heavy elements and dust essential to the formation of planets and life, at the same time as infall from the CGM adds gas with generally lower heavy element abundances. In this complex, turbulent, and dynamic environment, clouds of dense molecular gas are produced that are the sites of star formation. Within these molecular clouds, dense cores form that eventually gravitationally collapse and often fragment further to form stars with a wide range of masses. At least some of the angular momentum of the parent cores is retained during gravitational collapse, resulting in the formation of rotating circumstellar disks where planets form. The scope of this report of the Panel on the Interstellar Medium and Star and Planet Formation spans these widely disparate phases and structures over huge ranges of scale that are nevertheless inextricably linked.

Major advances in our understanding of the ISM and star and planet formation have been made over the past decade, encompassing the star-forming activity in nearby galaxies, the structure and properties of gas and dust in the local ISM, the fragmentation and collapse of dense gas to form stars and circumstellar disks, and the properties of those disks that seed the formation of planetary systems. To provide context for our recommendations in the coming decade, here we briefly outline some of the significant progress in these fields.

Studies of the galactic ISM have provided much more detailed characterizations of structures in both the dense and diffuse medium over the past decade. Far-infrared imaging and spectroscopy from Herschel emphasized that filamentary structures are ubiquitous in dense molecular clouds (Figure F.1, left), while observations from the Planck spacecraft and the Galactic Arecibo L-band Feed Array HI Survey (GALFA) 21 cm survey at the Arecibo Observatory demonstrated that filamentary structure is also prevalent in the diffuse medium. The polarized sub-mm emission measured by Planck (Figure F.1, middle) showed that dust grains are efficiently aligned by magnetic fields, which exhibit coherent structure over large scales and provide important tests for models of ISM dynamics.

Significant progress was also made in characterizing the spatial distribution of gas and dust in the local ISM. Velocity-resolved observations of [CII] 158 μm emission with both the Herschel spacecraft and the Stratospheric Observatory for Infrared Astronomy (SOFIA) have helped quantify the significant fraction of molecular gas that is not traced by CO emission (“CO-dark” gas). Large-scale multiband stellar surveys from the Sloan Digital Sky Survey (SDSS) and Pan-STARRS, coupled with state-of-the-art statistical methods, have helped develop novel 3D models for the spatial distribution of dust and

regional variations in the reddening curve. In the near term, distances and stellar characterizations from future Gaia data releases will enhance the power of these methods.

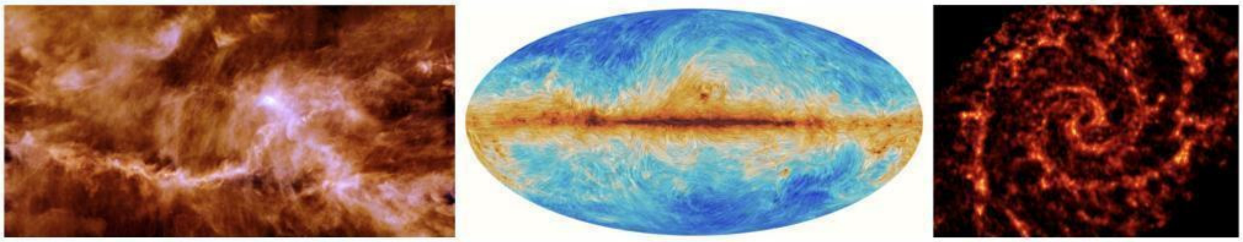


FIGURE F.1 (*left*) Composite far-infrared Herschel image of the B211/B213 filament in the Taurus molecular cloud. (*middle*) Full-sky Planck polarization map at 353 GHz. (*right*) CO map of giant molecular clouds in the core of the M74 galaxy. SOURCE: *Left:* https://sci.esa.int/s/8YYqz18, ESA/Herschel/PACS, SPIRE/Gould Belt survey Key Programme / Palmeirim et al., 2013, *A&A* 550, A38, 2013. *Middle:* Planck Collaboration, 2015, *A&A* 576, 104. ESA and the Planck Collaboration. *Right:* ALMA ESO/NAOJ/NRAO; NRAO/AUI/NSF, B. Saxton; Kreckel et al., 2018, *ApJL*, 863, L21. ALMA (ESO/NAOJ/NRAO); NRAO/AUI/NSF, B. Saxton.

On larger scales, the distribution of molecular gas in nearby galaxies has been mapped in exquisite detail using the Atacama Large Millimeter/Submillimeter Array (ALMA) (Figure F.1, right), showing how the properties of molecular clouds depend on the local environment. In particular, comparisons of ALMA CO maps to tracers of massive star formation on matched scales have revealed systematic variations in the star formation efficiency of molecular clouds, as well as the breakdown of star formation laws on cloud scales where the life cycles of star-forming clouds set by feedback or dynamics dominate. Observations with ALMA, Herschel, and optical integral field units (IFUs) have detailed the launching of outflows driven by feedback from massive star formation, revealing the cycling of ISM material into the circumgalactic medium (CGM) and the intergalactic medium (IGM) of galaxies.

Shifting to local studies of Milky Way molecular clouds, there have been substantial developments in our understanding of the many relevant physical scales and processes associated with star formation. Surveys of local star-forming regions with Spitzer provided the first complete censuses of their low-mass protostars and analyzed their spatial distributions relative to the cloud structures. Herschel observations provided the far-infrared data needed to determine protostellar luminosities and identify the youngest objects for further study (Figure F.2, left). Resolved sub-mm to cm continuum and spectral line imaging with ALMA and the Karl Jansky Very Large Array (JVLA) detected multiple protostars interacting with their disks while still embedded in their natal envelopes (Figure F.2, middle). Measurements of protostellar disk rotation and the kinematics of infalling envelopes from ALMA spectral line studies at high spatial resolution enabled measurements of the central masses, critical data for testing theories of protostar formation.

Parallaxes from Gaia and very long baseline interferometry (VLBI) have substantially improved estimates of young star luminosities and thereby ages, an essential aspect of characterizing the star formation histories of molecular clouds as well as for establishing protoplanetary disk lifetimes. Additional measurements of precise kinematics of young stellar populations from Gaia have increased the number of known nearby clusters and moving groups, facilitating more robust investigations of their initial mass functions, chemical homogeneity, and multiplicity statistics as a function of age. Photometric monitoring programs, at modest cadence from the ground (e.g., the Palomar Transient Factory [PTF]) and high cadence from space (e.g., the Microvariability and Oscillations of Stars telescope, Kepler), have revealed diverse protostar and pre-main sequence variability that signals ubiquitous, complex, and variable accretion phenomena.

Complementing this observational progress, computational modeling of star formation has advanced considerably over the past decade. Three-dimensional, adaptive resolution simulations

including radiative transfer, magnetohydrodynamics (MHD), and other physics (accretion, outflows) are now routine (if still expensive). Radiative transfer post-processing has also enabled more direct comparisons between the models and data. In many instances, these improvements have highlighted shortcomings in current models of important physical processes, necessitating ongoing study.

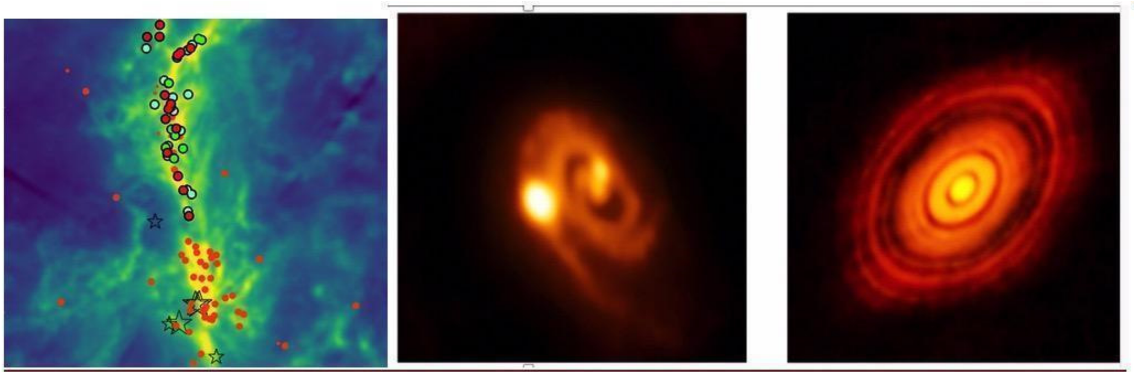


FIGURE F.2 (*left*): Column density map in the region of the Orion Nebula derived from Herschel with the positions of protostars superimposed. (*middle*): ALMA 1.3 mm image of dynamical interactions in the L1448-IRS3B protostellar binary system. (*right*): ALMA 1 mm image of narrow gaps and rings in the HL Tau circumstellar disk, perhaps sculpted by a young planetary system. SOURCE: *Left:* Megeath et al., *Science* White paper. Courtesy of S. T. Megeath et al., 2019, arXiv: 1903.08116. Reproduced with permission. *Middle:* ALMA/ESO/NAOJ/NRAO/J.J. Tobin; Tobin et al., 2016, *Nature*, 538, 483. *Right:* ALMA/ESO/NAOJ/NRAO; ALMA Partnership et al., 2015, *ApJL*, 808, L3. ALMA (ESO/NAOJ/NRAO).

On still smaller scales, new insights have emerged from high-resolution imaging of circumstellar disks (especially with ALMA), including constraints on the physical conditions, dynamics, and chemical abundances in the environments where planets form and grow, preliminary looks at demographic relationships between disk and host parameters, and the signatures of small-scale substructures in the spatial distributions of disk material. These latter features, in the forms of gaps, rings, spirals, and arcs on ~few astronomical unit (AU) scales, appear to be pervasive in early high-resolution samples, with major implications for models of planet formation and planet-disk dynamical interactions (Figure F.2, right). Many additional insights into planet formation have become available, with a particular highlight being the comprehensive survey of short-period exoplanets (primarily from Kepler), and their initial characterization. Unique information has also been gleaned on the nature of Jupiter’s core from the Juno mission, while New Horizons has provided the first exploration of a possibly primordial body in the Kuiper Belt.

On the theoretical side, novel ideas have emerged to better explain the transport of material and formation of solid bodies in disks during the planet formation epoch, including wind-driven accretion and accelerated planetary core growth via pebble accretion. Computational advances now allow simulations of these and other key processes from first principles, which can be compared against astrophysical data or in situ measurements. Laboratory experiments, including microgravity studies of dust particle collisions, chemical measurements of ice mantle formation and sublimation, and analyses of meteorites, provide critical inputs for both theoretical models and the interpretation of astronomical data.

Building on these many achievements over the past decade, this appendix discusses the opportunities for making further progress in characterizing and understanding the state of the ISM in the Milky Way and nearby galaxies, star formation, and planet formation, greatly assisted and informed by the contributions of more than 150 science white papers from the broader community.

Below, four key science questions and one discovery area identified by this panel are discussed and are listed for convenience near the end of this appendix in Box F.1. A summary of relevant facilities and enabling capabilities needed to answer these questions appears at the end in Table F.1.

F-Q1. HOW DO STAR-FORMING STRUCTURES ARISE FROM AND INTERACT WITH, THE DIFFUSE INTERSTELLAR MEDIUM?

While star formation occurs in molecular clouds, there remains considerable uncertainty on the mechanisms that control the formation and evolution of those clouds in the Milky Way and other galaxies. Interstellar gas must cool and condense over many orders of magnitude to reach the densities and temperatures necessary for star formation. The transitions in density, temperature, and chemical state are driven by turbulence, stellar feedback, nonequilibrium chemistry, and dust evolution. Disentangling these processes and characterizing their dependence on galactic environment in the Milky Way and beyond are critical challenges for the next decade.

Over the coming decade, observations will be capable of delineating the 3D structure of the gas, dust, and magnetic field, providing a new and dramatically detailed view of the dynamic ISM: "ground truth" to test theoretical models. Observations of other galaxies have less spatial detail, but will allow us to probe the star-forming ISM under different conditions, and over a range of metallicities. Milky Way observations will provide complementary data, including 21 cm emission, CO emission, dust extinction, interstellar absorption lines, starlight polarization, and polarized sub-mm emission from dust. These huge data sets will be a challenging opportunity for Big Data methodologies—for example, automated analysis pipelines run "locally" at the site of data acquisition that can distill multiterabyte datacubes into manageable data products, obviating the need for massive transfers and immediate human intervention.

F-Q1a. What Sets the Density, Temperature, and Magnetic Structure of the Diffuse ISM, Enabling the Formation of Molecular Clouds?

To understand the origin of star-forming clouds, we first must understand the state of diffuse gas in the Milky Way (MW) and other galaxies. The neutral gas can be characterized using emission and absorption in the 21 cm line to derive the HI spin temperature, revealing the cold neutral medium (CNM) and warm neutral medium (WNM) mass fractions, and the temperature in each phase. To constrain how the CNM/WNM fraction depends on galactic environment, HI absorption measurements in nearby galaxies are required. Imaging cold neutral gas structures in emission is also critical to uncover the cold gas dynamics, organization, and connection to star formation.

Our view of the Milky Way's ISM has been hampered by seeing it in projection. As a result, we have little information about volume density; observed quantities are line-of-sight averages; constraints on kinematics are highly incomplete; and studies of the 3D magnetic field are compromised. We are on the verge of a revolution, enabled by Big Data methods to exploit the stellar distances provided by Gaia, to construct a backbone for a 3D view of the ISM. Photometric and spectroscopic surveys plus Gaia distances already have been used to create spatial maps of dust extinction. High-resolution optical spectroscopy of absorption lines (NaI, KI, CaII, CH, CN, C_2) toward stars of known distances would allow the gas to also be dissected in 3D; 21 cm emission components can then be associated with optical absorption lines at the same velocity. Large surveys of stellar polarization (e.g., PASIPHAE) and filamentary HI features can outline the spatial structure of the magnetic field. Early results from these approaches are spectacular. For example, the 3D structure of the Orion A molecular cloud has been found to differ greatly from its projected structure. The first generation of 3D dust maps also show spatial variations of the extinction curve across the MW, revealing regional evolution of dust properties.

F-Q1b. How Do Molecular Clouds Form from, and Interact with, Their Environment?

Studying the formation of molecular clouds requires observations of regions where dramatic changes in temperature, density, and chemistry are occurring. In particular, it is critical to track the "CO-dark" gas, where CO is underabundant. UV observations from space offer a uniquely powerful window

into this process, by directly observing the absorption lines from HI, H_2, CO, CII, and many other molecules, atoms, and ions. With high spectral resolution, the dynamics and physical state of the gas can be characterized. Building on the existing framework of stellar distances from Gaia, UV spectroscopy toward stars with known distances can provide a 3D view of MW molecular cloud formation.

While UV spectroscopy can provide detailed information, the reach of such observations is limited to conditions where individual stars can be resolved (e.g., the Local Group with current facilities). The fine structure transitions of various atoms and ions provide means to trace gas in a variety of phases both in the Milky Way and nearby galaxies. A benefit of the far-IR lines, such as [CII] 158 μm and [OI] 63 μm (Figure F.3) is that they are relatively easy to excite and thus easily detectable even out to high redshifts with ALMA. A challenge is that they represent the contributions from gas in many different physical states. Enabling the full diagnostic potential of these lines requires high-velocity resolution far-IR observations coupled with matched resolution diagnostics of the ionized, atomic, and molecular gas phases (e.g., Hα, HI, and CO) to separate the emission arising from various phases. First, associating components of [CII] with HI can probe the thermal pressure in the diffuse ISM, a critical ingredient that mediates the phase transition between CNM and WNM. Second, the line intensities provide information on the heating of the gas, controlled by the photoelectric effect from small dust grains. Last, any components *not* associated with Hα, HI, or CO may arise from gas that is in transition between phases, most importantly the "CO-dark" H_2, of critical importance for measuring total molecular gas content especially at low metallicity.

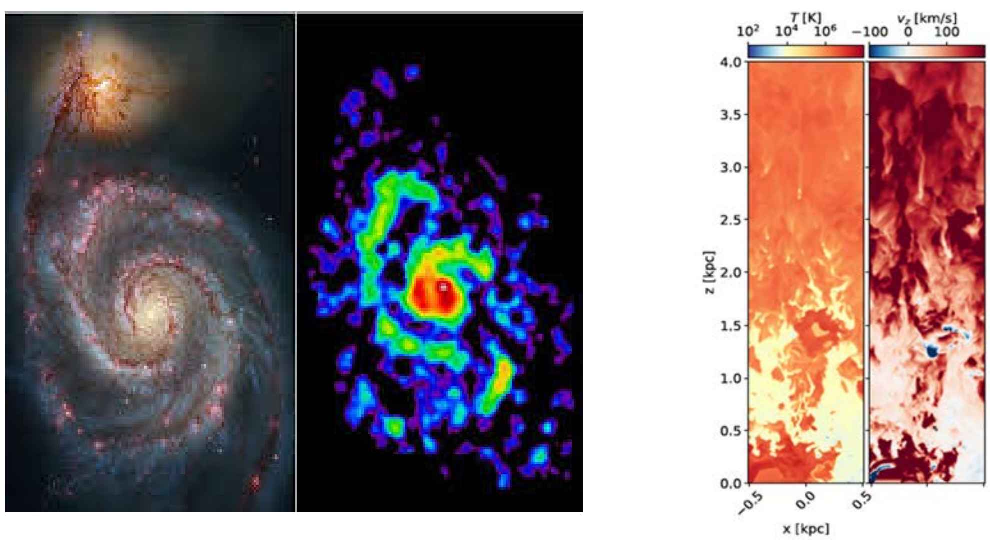


FIGURE F.3 (*left*) A Hubble Space Telescope image of M51. (*middle)* The [CII] 158 micron image of M51 obtained with SOFIA. (*right*) TIGRESS simulations of multiphase, turbulent, magnetized ISM with star formation and SN feedback. (Temperature and velocity are shown during an outflow-dominated period.) SOURCE: *Left:* NASA, Hubble Heritage Team, (STScI/AURA), ESA, S. Beckwith (STScI). Additional Processing: Robert Gendler. SOURCE: J.L. Pineda et al. 2018 ApJL, 839, L30; C. Fischer/DSI. *Right:* Adapted from C.-G. Kim and E.C. Ostriker, 2018, "Numerical Simulations of Multiphase Winds and Fountains from Star-forming Galactic Disks. I. Solar Neighborhood TIGRESS Model," *The Astrophysical Journal*, 853 173. © AAS. Reproduced with permission. doi:10.3847/1538-4357/aaa5ff.

Dust is critical to all stages of the transition between diffuse to molecular gas, through its roles in thermal balance, chemistry, and shielding. Our current understanding shows that the size distribution, composition, and overall abundance of dust relative to gas (the dust-to-gas ratio) change within the ISM. As these properties vary, the efficacy of the dust in its ISM roles is modified. Over the next decade, we need to develop a comprehensive picture of how dust evolves in the ISM, both in the Milky Way and

extragalactic environments. This will require observational constraints as well as theoretical models and simulations that track the dust life cycle. High-resolution UV spectroscopy can provide measurements of the depletion of heavy elements in the MW and nearby galaxies; low-resolution UV through mid-IR spectroscopy can trace the extinction curve; and mid-IR spectroscopy with JWST will better characterize the silicate absorption bands and other features. X rays can provide detailed information on dust mineralogy and dust-to-gas ratio, via measurements of absorption edges and scattering halos. Last, laboratory studies of candidate materials at all wavelengths are vital.

F-Q1c. How Does Injection of Energy, Momentum, and Metals from Stars ("Stellar Feedback") Drive the Circulation of Matter Between Phases of the ISM and CGM?

The ISM is largely driven by energy from stars. Over the next decade, it is critical to perform increasingly realistic MHD simulations that include transitions between ionized, atomic, and molecular gas phases, dust evolution, cosmic rays, and the inflow and outflow to the CGM (an example is shown in Figure F.1). Simulations on many different scales are needed, from galaxy scales (cell sizes of tens of pc) to molecular cloud scale (where the cell size may be below 0.1 pc). Including stellar feedback in the form of injection of energy and momentum as well as metal enrichment by nucleosynthetic products is required.

Enabling Capabilities

To characterize the neutral gas in nearby galaxies, HI 21 cm absorption and emission observations on cloud scales ~1" (~100 pc for D <= 20 Mpc) with improved sensitivity are needed (this requires ~10× the collecting area of the JVLA). To take full advantage of the stellar parallaxes from Gaia (or elsewhere) to map the 3D structure of the Milky Way's ISM, high resolution (R > 60,000) optical spectroscopy, with sensitivity to obtain spectra for all V < 15 O, B, and A stars of known distances, will enable revolutionary studies of the temperatures, densities, and kinematics of the diffuse gas in the Milky Way. Pilot studies of feasibility for using a larger sample of F, G, and K stars for absorption line studies are crucial to densely sample structures on cloud scales. High-resolution ($R > 10^5$) spectroscopy (to resolve line profiles of absorption from cold gas) in the UV (~1000–3000 Å) toward stars with significant extinction is needed to probe the formation of molecular gas. This requires effective apertures ~3–5× present HST/COS capabilities. Similarly, expanded starlight polarimetry surveys to map the ISM magnetic field toward stars with known distances are needed to map out the structure of the Milky Way's magnetic field.

Wide-field (up to ~10 deg^2 for MW regions), high-sensitivity mapping of velocity-resolved (~0.1 km/s) far-IR lines ([CII] 158 µm, [OI] 63 µm, and others) in the MW and nearby galaxies is needed to study the "CO-dark" gas, along with matched resolution HI, CO, and Hα observations from existing facilities. Observational (X ray to mm), theoretical, and laboratory studies are important for better characterization of the properties and evolution of dust, and its role in ISM thermal and ionization balance. Last, making full use of the proposed expanded set of observations will require galaxy simulations including realistic feedback, multiphase gas, radiative transfer, CGM/ISM inflow and outflow, cosmic ray acceleration and transport, and a live dust model.

F-Q2: WHAT REGULATES THE STRUCTURE AND MOTIONS WITHIN MOLECULAR CLOUDS?

Molecular clouds are structurally complex, with substructure arising from MHD turbulence, chemistry, and self-gravity. The processes that drive the turbulence remain unclear. Stellar feedback, in

the form of flows external to the cloud driven by expanding HII regions and supernova blastwaves, is thought to play an important role. Within star-forming clouds, energy and momentum are also injected by protostellar outflows, and cosmic rays play key roles in heating the gas. Last, the role of gravity in driving supersonic motions cannot be neglected.

F-Q2a. What Processes Are Responsible for the Observed Velocity Fields in Molecular Clouds?

Large-scale motions within molecular clouds are observed to be highly supersonic. The kinematics (in projection) are currently studied using cold gas tracers such as CO 1-0 or 2-1. Maps of higher-excitation far-infrared lines—such as [CII] 158 µm, [OI] 63 µm, rotational transitions of OH, CH^+ and other hydride ions, and high-J rotational transitions of CO—would be invaluable to trace gas heating by turbulent dissipation. With high spectral resolution, such maps will reveal the kinematics of the warm gas. To clarify the role of the magnetic field in turbulence and cloud structure, we need maps of polarized dust emission to reveal the geometry, and Zeeman effect measurements of field strength using species such as CN and OH. Also needed are MHD simulations, on various length scales, that include realistic gas physics, including chemistry and line emission.

F-Q2b. What Is the Origin and Prevalence of High-Density Structures in Molecular Clouds and What Role Do They Play in Star Formation?

The densest gas, in which stars form, generally comprises only a few percent of the total cloud mass, leading to low global star formation efficiencies. An observational census of the dense gas as a function of interstellar environment, and understanding how dense structures form and evolve, has important implications for understanding galactic-scale star formation. Moreover, dense structures set the initial conditions for subsequent collapse to stars and disks.

In nearby molecular clouds, the densest gas often appears filamentary. The pervasiveness of filaments in dust continuum images is one of the key results from Herschel. While filaments likely dominate the mass budget of the dense molecular gas where stars form, the understanding of their formation, fragmentation, as well as the degree to which they contain sub-structure remain controversial. Furthermore, it is not clear whether filaments are a widespread and critical step in star formation across galaxies of different properties. The key way to study filament formation, growth, and dispersal is via dense-gas kinematics using molecular species like ammonia, N_2H^+, deuterated molecules, and CO isotopologues (an example is shown in Figure F.4). High-resolution observations can untangle and measure the gas flows within molecular clouds that assemble filaments, and search for infall motions and velocity oscillations along filaments that lead to core formation. While ALMA and the Green Bank Telescope are making important strides in this area, the progress is slow owing to the limited mapping speed for sensitive multitracer observations (e.g., lines tracing lower-density gas, high-density gas, and shocks). In tandem, for a large sample of star-forming clouds, we need high spatial resolution (<0.1 pc) mapping of the magnetic field in the filaments and cores to constrain field geometry and the extent to which magnetic fields provide support against gravitational collapse (Figure F.4). While ALMA and SOFIA are making important strides in this direction, larger samples probing diverse interstellar environments are essential.

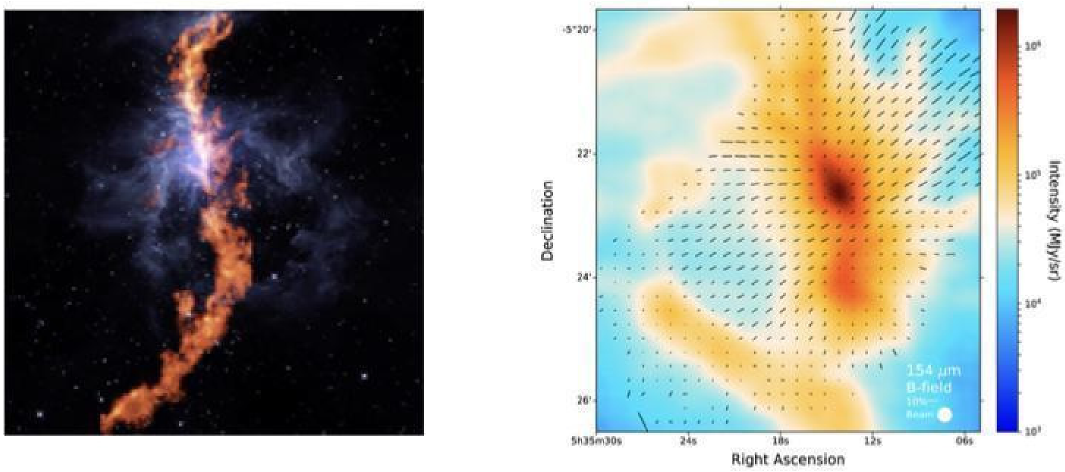


FIGURE F.4 (*left*) A ribbon of ammonia—a tracer of star-forming gas—in the Orion Nebula seen with the Green Bank Telescope. *(right*) Orion 154 µm polarization from the SOFIA HAWC+ polarimeter. The polarization vectors were rotated by 90 degrees to indicate the direction of the magnetic field projected in the sky plane. SOURCE: Kirk et al., 2017, *ApJ*, 846, 144. GBP/AUI/NSF. *Right:* Adapted from D.T. Chuss et al 2019, "HAWC+/SOFIA Multiwavelength Polarimetric Observations of OMC-1," *The Astrophysical Journal*, 872 187. © AAS. Reproduced with permission. doi:10.3847/1538-4357/aafd37

In external galaxies, detailed studies of dense gas have been limited to only the closest galaxies owing to the faintness of the relevant lines. While we cannot resolve individual dense structures, systematic measurements of the physical state of the cold and dense ISM across the full range of galactic conditions and environments found in the local universe will uncover processes that regulate the fraction of the star-formation feedstock—dense molecular gas. Density, excitation, and chemistry play central roles in many theories of star formation; linking these quantities to galaxy structure and galaxy evolution across redshift requires multitransition, multispecies (HCN, HCO^+, CO isotopologues, different excitation lines) mapping of a diverse set of galaxies.

F-Q2c. What Generates the Observed Chemical Complexity of Molecular Gas?

To fully interpret observations of molecular gas, we need to invest in testable chemical theories that can explain observed chemical abundances. Current astrochemical models fail to fully explain the complexities of observed molecular abundance ratios over a range of densities. In addition, observational efforts are needed to place rigorous constraints on models of chemical evolution at each stage of the star and planet formation process. In particular, understanding the formation pathways and excitation of large astronomical molecules, all the way to (pre)biotic molecules such as glycine and glyceraldehyde (the simplest sugar), and connecting molecule formation with processes happening in icy mantles of dust grains is critical. Laboratory studies are vital for measuring and quantifying critical pathways within astrochemical reaction networks.

Enabling Capabilities

Far-IR/sub-mm line imaging, ideally with high spectral resolution (R > 5e4), is needed to distinguish heating by turbulent dissipation, stellar radiation and/or cosmic rays. To address the role of magnetic fields in cloud structure and dynamics, high-resolution (<0.1 pc, ~10" to resolve filaments in MW clouds) maps of polarized dust emission are needed along with maps of circularly polarized emission

from CN with <0.1 pc (~10") resolution and sensitivity to Stokes V < 0.5 mK at 0.2 km/s resolution. MHD simulations with chemistry, covering a range of scales from molecular cloud environments to dense cores, are essential to interpret observations by providing synthetic spectra and magnetic field maps for statistical comparisons.

High spatial (<0.1 pc) and velocity (<0.1 km/s) resolution radio and mm lines tracing gas at a variety of densities (radio: N_2H^+, NH3, deuterated molecules, mm: N_2H^+, CO isotopologues) for MW clouds are needed to study filaments. Far-IR and sub-mm polarization maps are important to trace magnetic field structure on the scale of molecular clouds and their substructures (<0.1 pc).

Cloud-scale (~100 pc, <1" at D = 20 Mpc) radio and mm spectroscopy of large samples of nearby galaxies to detect HCN, HCO^+, CO isotopologues, and different excitation lines (>10× fainter than CO) are crucial to measuring the physical state of dense gas, then relate that to environment and the star formation efficiency.

Laboratory measurements and modeling of ices, gas phase chemistry, dust surface reactions, and the spectra of complex molecules are essential to understanding dense gas phases. Deep (RMS noise of 100 μJy/beam) cm/mm line surveys with spectral resolution to overcome line confusion (<0.1 km/s) and detect complex molecules with the angular resolution ~1" are required to isolate local environmental conditions.

F-Q3: HOW DOES GAS FLOW FROM PARSEC SCALES DOWN TO PROTOSTARS AND THEIR DISKS?

The fundamental challenge of star formation is to understand how processes spanning an enormous range of scales, from parsec-scale turbulent flows to sub-AU disk accretion, combine to produce the apparently universal stellar initial mass function (IMF) and planet-forming disks. Existing or nearly completed facilities, if complemented by some crucial future investments, will enable substantial progress in our understanding of critical fragmentation and accretion processes in the coming decade.

F-Q3a. How Do Dense Molecular Cloud Cores Collapse to Form Protostars and Their Disks?

The initial phase of star formation is controlled by the collapse and fragmentation of dense molecular gas onto a disk, which then transports mass inward and angular momentum outward. This process is dictated by the density structures of dense molecular gas "cores," which are yet to be well resolved observationally. Moreover, the lifetimes of these cores are unknown: Do they represent a fixed mass reservoir from which bound stellar systems accrete, or do they evolve over the collapse time scale? Observations at (sub-)mm to cm wavelengths at the highest resolution available (~500 AU) for a small sample of cores reveal that they diverge from simple assumptions of spherical symmetry and solid body rotation (Figure F.5).

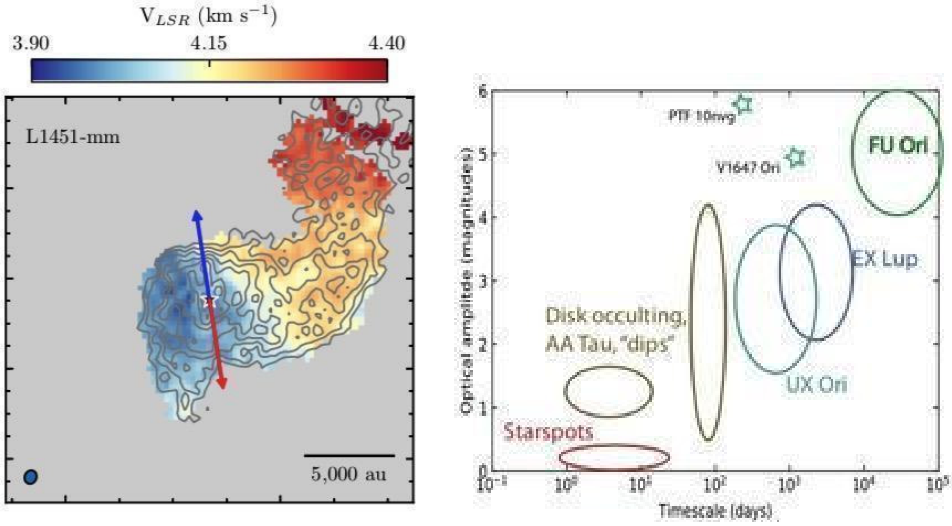


FIGURE F.5 (*left*) JVLA observations of the protostellar core L1151-mm in the NH_3 (1,1) line reveal a velocity field suggesting an angular momentum distribution with contributions from solid-body rotation and infall motions. The position of the embedded protostar is indicated by the star, and the outflow direction by the arrows. (*right*) The schematic parameter space of (optical) young stellar variability occupies a wide range in both time scale and luminosity. All-sky surveys with spectroscopic follow-up will enable critical characterization of these accretion phenomena. SOURCE: *Left:* Adapted from Jaime E. Pineda et al 2019, "The Specific Angular Momentum Radial Profile in Dense Cores: Improved Initial Conditions for Disk Formation," *The Astrophysical Journal*, 882 103. © AAS. Reproduced with permission. doi:10.3847/1538-4357/ab2cd1. *Right:* Modified from Hillenbrand and Findeisen, 2015, *ApJ*, 808, 68.

Addressing these fundamental questions demands more complex modeling, and that will require measurements of the internal kinematics and density fluctuations on scales below 1000 AU for a much larger sample of cores. Insights on how magnetic fields mediate the collapse process can also be obtained from measurements of field morphologies—for example, from polarized (sub-)mm to cm dust emission and Zeeman splitting from key molecular line tracers. With such refined observational constraints, we can assess core lifetimes, their susceptibility to fragmentation and binary star formation, and ultimately the link between core masses and the stellar IMF.

F-Q3b. How Do Protostars Accrete from Envelopes and Disks, and What Does This Imply for Protoplanetary Disk Transport and Structure?

The large-scale mass transfer intrinsic to core collapse continues down to smaller scales, with material transported through the disk and onto the central star. A better understanding of these disk transport processes (see also section F-Q4, below) is an important complement to the detailed studies of the core structures outlined above. One compelling avenue for new insights on these processes comes from time-domain measurements, tracking broadband and spectroscopic variability on a range of time scales.

Figure F.5 illustrates the wide range of variability that should be characterized. High cadence (~seconds to minutes) targeted photometric surveys will identify low-amplitude, rapid variability in the inner disk and on the protostellar surface (flares) up to time scales of hours to days owing to a host of processes, from quasi-regular accretion via disk-magnetosphere interaction to (still poorly understood) extinction events. The latter may be produced by MHD turbulence or even signal the presence of perturbing bodies (e.g., planets). In contrast, deep, large-area surveys (e.g., the Vera Rubin Observatory) will provide complementary data on rarer, high-amplitude variability like FU Orionis star outbursts,

which dramatically affect both the masses and structures of stars and their disks. Because these outbursts are infrequent and long-lived (decades), samples of >10^5 young stars are necessary to constrain their properties and occurrence rates (i.e., a census of regions stretching beyond ~2 kpc). Furthermore, Spitzer and Kepler monitoring campaigns showed that low-amplitude, medium-duration (~month) variability is common; there is a vast parameter space of unexplored time-domain behavior poised for discovery. On any time scale, a comprehensive characterization of time-variable accretion activity requires both photometric and spectroscopic studies of large samples, spanning a range of masses and evolutionary states. Near-simultaneous spectroscopic measurements at optical to near-IR wavelengths (ideally to ~5 μm) are needed to determine accretion rates and accurate stellar parameters (mass, radius, effective temperature).

F-Q3c. Is the Stellar Initial Mass Function Universal?

The form of the stellar initial mass function (IMF), provides a fundamental test of star formation theories and is crucial to the interpretation of the composite spectra of distant galaxies. Whether the IMF is universal or depends upon the star-forming environment is not a settled question. To characterize mass functions in extremely dense, massive regions requires the study of clusters—in which most stars form—and associations nearer the galactic center and in nearby galaxies. The Magellanic Clouds offer an especially compelling opportunity to probe the effects of metallicity, which has been shown in local samples to have an impact on stellar multiplicity. To reach these distant environments requires better spatial resolution to deal with stellar crowding, and improved sensitivity to spectroscopically probe the peak of the stellar mass function, ideally to masses <= 0.2 solar masses, or K magnitudes ~ 23 at 50 kpc, with resolution < 0.004 arcsec.

Enabling Capabilities

Measurements of the spatial distributions of physical conditions (densities, temperatures) and kinematics (rotation, collapse, and turbulent motions) in star-forming cores require sensitive, high-resolution mapping capabilities of key molecular gas and dust tracers accessible from (sub-)mm to cm wavelengths. Rotational transitions of CO isotopologues, ammonia, simple molecules (CN, HCN), and chemically important ions (N_2H^+) are crucial diagnostics, along with very deep dust continuum data at a range of wavelengths and with full polarization capabilities. Observations at the longest wavelengths (~few cm) are especially important for probing high-mass star formation in environments with very high dust extinction, as well as better penetration of high optical depths in the densest parts (generally smallest-scales) of infalling low-mass cores. Those observations need sub-arcsecond spatial resolution (< 1000 AU for massive star-forming cores at ~kpc distances) and velocity resolutions of ~50–100 m/s in the spectral lines (to constrain protostar masses, subsonic turbulent line widths, and small collapse motions) at ~10× improved sensitivity compared to current facilities (e.g., the JVLA) to assemble measurements for a sufficiently large sample in diverse environments.

Characterizing young variable objects requires high signal-to-noise ($\gtrsim 50$) optical and near-IR spectroscopy with R $\gtrsim$ 20,000 on $\gtrsim$8 m telescopes, to identify faint protostellar absorption features (diminished by veiling continuum emission), disentangle photospheric emission from hot accretion continua and lines, measure nonthermal line profiles, and uncover distant, heavily extincted, star formation. To optimally synchronize these capabilities with time domain surveys will also require a sophisticated alert system to trigger and prioritize follow-up. The ANTARES software "instrument" is an excellent example of how the community might achieve such goals.[1] Open-source platforms enable broad

[1] A. Saha, Z. Wang, T. Matheson, G. Narayan, R. Snodgrass, J. Kececioglu, C. Scheidegger, et al., 2016, ANTARES: Progress towards building a 'broker' of time-domain alerts, *Proceedings of SPIE 2016: Observatory*

swaths of the community to develop unique alert brokers for a range of systems. Collaboration with computer scientists will also be key to training brokers to find new young star phenomena that were too rare or faint for previous discovery.

The requisite deep photometry in a narrow-field, moderate-resolution (R ~ 4000) spectroscopy, and precise astrometry to distinguish cluster membership, can be done only with AO-equipped ELTs with enough sensitivity to at least reach the peak of the IMF at ~0.2 $M_\odot$ in the Magellanic Clouds ($K \gtrsim 23$). JWST and WFIRST can provide wider field IR imaging to study outer cluster regions, crucial to address mass segregation. Wide-field optical/NIR imaging and spectroscopy of many nearby low-mass star-forming regions might in sum provide a test of the low-density IMF.

F-Q4: IS PLANET FORMATION FAST OR SLOW?

A robust determination of the epoch of planet formation is a challenge for the coming decade. Are planets made slowly, over a time scale spanning their natal disk lifetimes? Or do they form quickly, at an early phase that overlaps with the epoch of star and disk formation? The former is the classical theoretical prediction, and suggests that disk properties should be interpreted as the initial conditions for planet formation. The latter implies instead that those properties may already be modified by dynamical interactions with young planetary systems. Building on recent advances, a direct route to answering this question focuses on the origins and demographics of *substructures* in circumstellar disks. That work needs to be supplemented with efforts to advance our capabilities to quantify the physical and chemical conditions of the disk material. Progress in these areas will facilitate an improved understanding of angular momentum transport in the disk and accretion onto the star, and particularly how those processes are related to the turbulence that impacts many aspects of planetary growth.

F-Q4a. What Are the Origins and Demographics of Disk Substructures?

High-resolution observations at near-IR and (sub-)mm wavelengths have revealed ubiquitous, small (~few AU) perturbations to disk structures in the forms of rings, gaps, spirals, arcs, and other asymmetries. As an illustration, Figure F.6 shows a gallery of disks imaged at high resolution in the mm continuum. These disk substructures are likely to be the hallmarks of localized particle concentrations at gas pressure maxima. Pressure maxima in disks can be produced by various (magneto)hydrodynamic instabilities, and in this interpretation the observed substructures are fluid mechanical features that may catalyze the planet formation process. Alternatively, pressure maxima and substructures could be the signposts of dynamical interactions with an already-formed generation of planets. That hypothesis would falsify the classical planet formation models that predict very slow time scales in the outer regions of disks, and point instead to more efficient mechanisms to assemble giant planets with (at least initially) long orbital periods (perhaps employing rapid planetesimal formation and pebble accretion).

The physical mechanisms at play in both of these possible interpretations are broadly understood, but substantial computational work is required to make quantitative predictions that capture the interaction of magnetohydrodynamic, thermal, and chemical effects in disks. The key observational clues on the origins of disk substructures are expected to come from detailed characterizations of their morphologies (e.g., gap shapes), kinematics (non-Keplerian deviations), and physical conditions (density contrasts, pressure gradients, particle size distributions, etc.). Enhanced spatial resolution, particularly at more transparent continuum wavelengths (i.e., the cm radio bands) and in molecular spectral line emission, are essential, both to resolve the larger substructures and to probe the currently inaccessible features in the inner few AU of nearby disks. Ultimately, the information gleaned from these detailed

---

Operations: Strategies, Processes, and Systems VI (A.B. Peck, R.L. Seaman, and C.R. Benn, eds.), Vol. 9910, International Society for Optical Engineering (SPIE), Bellingham, Wash.

characterizations will benefit substantially from demographic explorations of how substructure properties depend on external factors like the host mass, age, and environment.

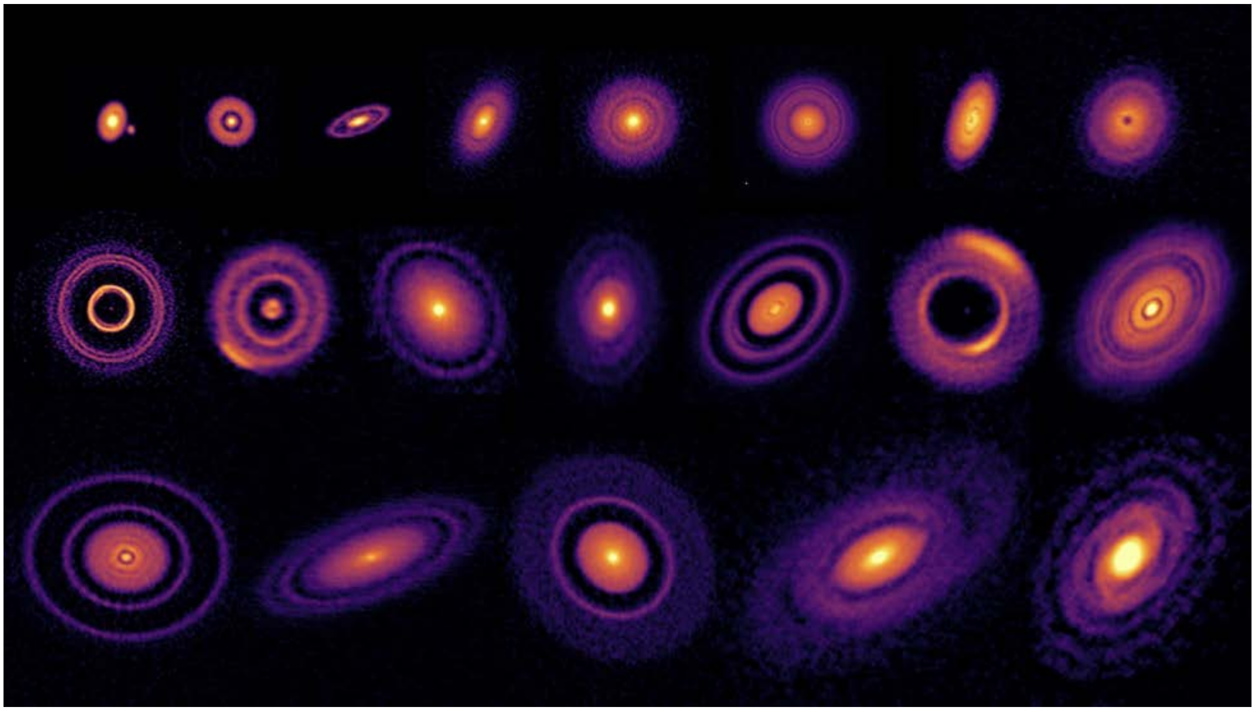


FIGURE F.6 A gallery of ~1 mm dust continuum images of nearby circumstellar disks made with the Atacama Large Millimeter/Submillimeter Array (ALMA). At sufficiently high resolution (30–50 mas, corresponding to a few AU, or better), disks are found to be riddled with fine-scale substructures that have a variety of morphological forms, sizes, amplitudes, and spatial distributions. SOURCE: Adapted from A. "Dust Unveils the Formation of a Mini-Neptune Planet in a Protoplanetary Ring," Sebastián Pérez et al, 2019, The Astronomical Journal, 158 15. Reproduced with permission. doi:10.3847/1538-3881/ab1f88. B. "The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, Sample, Calibration, and Overview," Sean M. Andrews et al 2018, The Astrophysical Journal Letters, 869 L41. © AAS. Reproduced with permission. doi:10.3847/2041-8213/aaf741. C. "The Eccentric Cavity, Triple Rings, Two-armed Spirals, and Double Clumps of the MWC 758 Disk," Ruobing Dong et al, 2018, The Astrophysical Journal, 860 124. Reproduced with permission. doi:10.3847/1538-4357/aac6cb. D. "CO and Dust Properties in the TW Hya Disk from High-resolution ALMA Observations," Jane Huang et al, 2018, The Astrophysical Journal, 852 122. © AAS. Reproduced with permission. doi:10.3847/1538-4357/aaa1e7. (E. "CO and Dust Properties in the TW Hya Disk from High-resolution ALMA Observations," ALMA Partnership et al 2015, The Astrophysical Journal Letters, 808 L3. © AAS. Reproduced with permission. doi:10.1088/2041-8205/808/1/L3.

F-Q4b. What Is the Range of Physical Environments Available for Planet Formation?

The metamorphosis of disk material into planetary systems involves a complex set of physical processes. Models for those processes are always limited by imperfect knowledge of the planet formation environment—the spatial variations of temperatures, densities (for gas and solids), chemical composition, particle sizes (and other microphysical properties), gas dynamics (turbulence), and magnetic fields. Over the past decade, considerable progress has been made on vetting the approaches available to constrain those environmental properties. In the next decade, we are poised to transition from that exploratory phase into a quantitative era that enables progress toward more predictive models of disk behavior.

The physical conditions in disks can be determined from a range of data sets and methodologies. The most promising options for determining the temperature, density, chemical, and dynamical structure

of the gas rely on spatially and spectrally resolved observations of molecular emission lines with a variety of optical depths and excitation states through the (sub-)mm bands and into the far-infrared (also, an important link to ices will be available in the mid-infrared with JWST). To anchor estimates of column densities onto an absolute scale, reference to measurements of a more direct gas mass tracer, particularly the isotopologue HD, would be beneficial. For the solids, the spatial variation in the particle densities and size distributions can be constrained with spatially resolved maps of the spectral shape and linear polarization of the (sub-)mm through cm continuum emission. Magnetic field strengths and geometries may be accessible through the Zeeman effect in various spectral features throughout the (sub-)mm bands.

F-Q4c. How Do Turbulence and Winds Influence the Evolution of Structure in Disks?

The paradigm that turbulence enables the angular momentum transport that drives accretion in disks has been challenged over the past decade. Angular momentum loss in a magnetohydrodynamic disk wind has emerged as an important alternative, which may be more compatible with (still limited) observations of minimal (nonthermal) spectral line broadening and the strong confinement of dust continuum features (both radially, in substructures, and vertically, in characteristic heights) in disks. Aside from better characterizing the different potential transport process, continued efforts to measure turbulence are highly desirable because of the diverse roles it plays in planet formation—from limiting the collisional growth of solids, to affecting the formation of planetesimals, and beyond to regulating the structures of gaps and the migration of protoplanets in gas disks.

Enabling Capabilities

The science issues of this question are best addressed with high-resolution measurements of the continuum and spectral line emission accessible to (sub-)mm/cm interferometers. Current facilities reach 20–70 mas resolution (3–10 au for nearby disks), with limited surface brightness sensitivity. Efforts to resolve larger substructures and discover new features that are smaller or located within ~10 AU of their host stars are essential, with a targeted factor of ~5 improvement in resolution (to ~5 mas, or sub-AU scales). Substantial sensitivity improvements will be necessary to probe the (~percent level) polarization and spectral line emission on those scales, as well as especially faint but chemically important emission lines at coarser resolution. An appropriate metric is an order of magnitude decrease in the noise level for a fixed integration compared to current facilities (e.g., JVLA, ALMA). High-resolution access to the cm/radio bands are crucial: the lower continuum optical depths there offer unique access to the properties of disk solids, particularly in the inner disk (terrestrial planet region), and measurements of free-free emission and H recombination lines that could be the key to quantifying disk winds. Those interferometric capabilities would strongly benefit from complementary observations of spectral lines throughout the far-infrared (particularly from water vapor and HD), as well as improvements in high-resolution images of disk structures in optical/infrared scattered light from dust grains entrained in the gas. The latter can be achieved with similar targets for improvements in resolution (~5× better) and sensitivity (~10× deeper) as indicated for the (sub-)mm/cm interferometers.

DISCOVERY AREA: DETECTING AND CHARACTERIZING FORMING PLANETS

Opportunities to characterize young planets and their circumplanetary material while they are still embedded in their natal circumstellar disks will lead to major advances in our understanding of the planetary accretion process, the formation of early atmospheres and satellites (moons), and the origins and evolution of both the solar system and the exoplanet population. Observations of dynamically cleared gaps in circumstellar disks may point to a population of young giant planets with masses ~ 10 $M_{\oplus}$ to 10

M_{Jup} and orbital separations of ~10–100 AU. At least one example of a system with actively accreting, young, giant planets has been found in a cleared disk gap, as shown in Figure F.7. Current observations with 8–10 m telescopes are sensitive to only the most massive planets, but extrapolations to ~30 m class facilities and reasonable improvements in instrumental capabilities can help find and characterize many more examples. In the longer term, the goal is to detect forming planets across a broad range of masses and orbital distances, transforming our understanding of planet formation and early planetary system evolution.

F-DA1. How Do Planets and Their Satellites Grow?

The transport of material from the circumstellar disk to a planet is mediated by a circumplanetary disk (CPD). By analogy to the circumstellar disk/host star system, we expect a variety of accretion diagnostics from the CPD/planet system, including Balmer line emission from magnetospheric funnel flows and excess UV/blue continuum from accretion shocks. These features have been observed in the few available CPD candidates associated with massive (> few M_{Jup}) planets (e.g., Figure F.7). Such measurements constrain the protoplanet accretion rate, informing models that describe how giant planet envelopes grow from flows across the gaps that planets sculpt in their parent disks. The physical conditions, spatial structures, and dynamics of CPDs are controlled by the combined gravitational potential of the protoplanet and host star, the local heating, and the mechanics of mass transfer from the circumstellar disk reservoir. Direct imaging measurements of the CPD spectral energy distribution (SED) would provide crucial insights on the thermal structure, and therefore constraints on the protoplanet luminosity and mass. Measurements of the CPD at (sub-)mm/cm wavelengths are sensitive to the CPD mass (in solids) and potentially its size, which scales with the planet's Hill radius (and thereby mass). Taken together, these measurements offer foundational boundary conditions for models of satellite formation. Ideally, they would be complemented with measurements of the CPD gas from molecular emission lines, as well as mm/cm-based estimates of the widths, depths, and kinematic perturbations of the associated gaps in the circumstellar disks (all sensitive to the protoplanet mass), to provide multiple diagnostics of CPDs and their planet hosts.

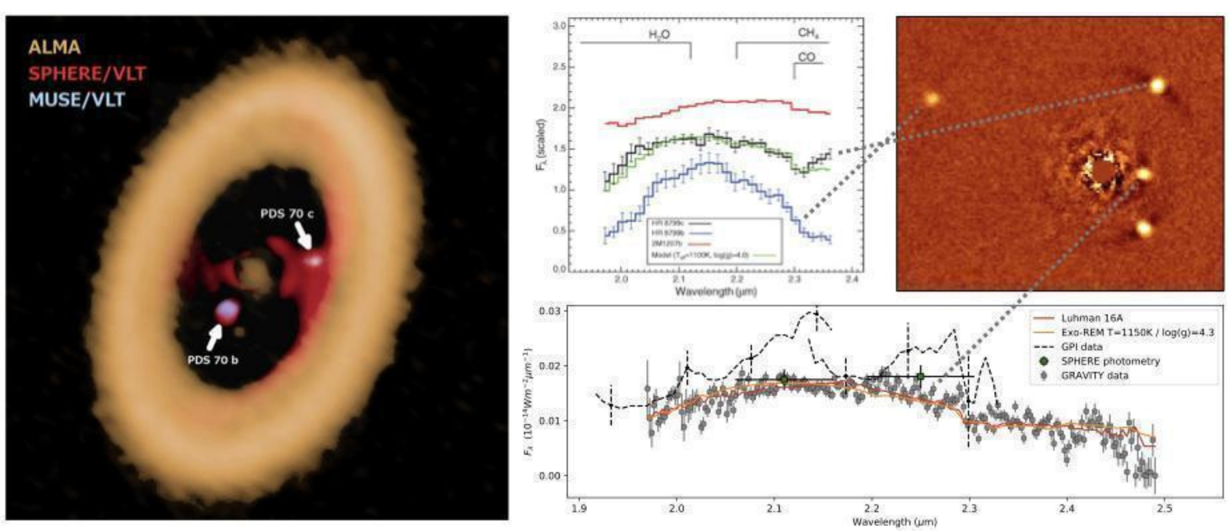


FIGURE F.7. (*left*) A composite image of the young PDS 70 system, showing a disk sculpted by two accreting (massive) giant planets with associated CPDs. (*right*) Low-resolution near-infrared spectroscopy from directly imaged planets in the prototype HR 8799 system. Considerable improvements in spectral resolution and imaging capabilities will enable measurements of molecular absorption features for lower mass planets, closer to their hosts. SOURCE: ALMA (ESO/NAOJ/NRAO), A. Isella; ESO. *Upper left spectrum:* Konopacky et al., 2013, *Science*, 339,

1398. *Bottom spectrum:* GRAVITY Collaboration, S. Lacour, et al., “First direct detection of an exoplanet by optical interferometry - Astrometry and K-band spectroscopy of HR 8799 e,” *A&A* 623 L11 (2019), doi:10.1051/0004-6361/201935253. *Image:* Adapted from A.J. Skemer, et al., "High contrast imaging at the LBT: the LEECH exoplanet imaging survey", Proceedings of SPIE 9148, Adaptive Optics Systems IV, 91480L (21 July 2014); doi:10.1117/12.2057277.

F-DA2. What Are the Atmospheres of Long-Period Giant Planets Like at Their Formation Epoch?

Direct imaging detections of young planets at wide orbital separations present compelling opportunities for spectroscopic characterization (e.g., see Figure F.7). The ability to measure absorption features from relatively extended planetary atmospheres would enable the first compositional analyses of planets at their formation epoch. These can be compared with measurements of their highly irradiated counterparts at small orbital separations (from transit spectroscopy) or their more evolved analogues around main-sequence stars (found via direct imaging) to better understand the chemical evolution and diversity of giant planet atmospheres. Moreover, estimates of atmosphere properties would help extrapolate a spectral sequence from young low-mass stars and brown dwarfs to offer some first-order benchmarks for key planetary properties (particularly masses). This would enable a rough empirical calibration of the link between the properties of planets and the disk gaps they sculpt, permitting a rubric to extrapolate toward the planet properties needed to explain more subtle disk features.

F-DA3. How Do the Orbital Architectures of Planetary Systems Evolve?

When an ensemble of protoplanets (and their CPDs) have been imaged, we can infer the giant planet mass function and its variation with orbital separation *at the formation epoch*. Then, by empirically characterizing how these planets perturb their disks as a function of their mass, we could extrapolate to indirectly infer constraints on the mass function into the super-Earth regime by measuring more subtle disk substructures. These young planetary architectures could then be compared with their more evolved counterparts around main-sequence hosts found from microlensing surveys and direct imaging campaigns. That comparison would provide novel constraints on the distribution of planetary migration histories. For example, if more long-period giant planets are identified at early times, it may indicate that the radial velocity and transit populations of exoplanets could have migrated from much more extended initial architectures.

Enabling Capabilities

The goals associated with this discovery area can be achieved with improved capabilities for deep direct imaging campaigns in the optical/IR, particularly when including modest resolution spectroscopy. Success will require deeper contrast limits (to search for long-period Saturn analogues around a range of host star masses requires near-IR contrast limits ~15× better than current ground-based 8–10 m telescope capabilities), improved adaptive optics systems for higher Strehl ratios to broaden the search for planets orbiting fainter hosts (e.g., R ~ 13–14 or better), smaller inner working angles to probe orbital separations of ~5–10 AU (~30–60 mas), and better resolution (~10 mas) to optimize contrast by resolving the associated disk gap and potentially the CPD. These goals are achievable from the ground with “extremely large” aperture telescopes (i.e., ~30 m diameter mirrors). There would be considerable value added with complementary instrumentation in the thermal infrared, ~3–5 microns (L and M bands), where planet and CPD contrasts are significantly improved. An instrument that can access the 10 and 20 μm windows (N and Q bands) would optimize the CPD contrasts for both the host star and surrounding disk. Moreover, assembling a more complete spectral energy distribution (SED), from optical to mid-infrared, will help disentangle bulk atmosphere properties of protoplanets and the thermal structure of CPDs. Spectroscopic

capabilities coupled to the adaptive optics system are also essential, both to probe atmosphere features and accretion diagnostics. Integral field units or analogous technology (including interferometric capabilities) with resolving power of at least a few hundred (and preferably up to a few thousand) are necessary. Links to the substructures these protoplanets sculpt in their natal disks will require the improved observational capabilities for (sub-)mm/cm interferometry data discussed in detail in the previous section.

THEORY AND MODELING

In tandem with the improved observational capabilities discussed earlier, progress will be further enabled by theoretical and modeling advances, which are enabling capabilities for all of the science questions. In common with other fields, computational progress in ISM, star formation, and planet formation problems requires ongoing improvements in the size and efficiency of large, multiscale numerical simulations that include MHD and self-gravity. The largest simulations—for example, of the launching of galactic winds owing to star-formation feedback into the ISM (F-Q1)—are at the limit of what is possible given existing supercomputers. Radiative transfer is required to create synthetic observations that can be compared against data, and in some problems as a dynamical actor (F-Q1, F-Q2, F-Q3, F-Q4, F-DA). Funding for the development of faster, more accurate, and more scalable algorithms in these areas is key, as it can improve the fidelity of simulations more rapidly than is possible from iterative advances in machine size (F-Q1, F-Q2, F-Q3, F-Q4). Also common with other fields is the likelihood that, over the next decade, a greatly increased fraction of compute resources will be devoted to machine learning (F-Q1, F-Q2, F-Q3). Machine learning has multiple potential applications, including the identification of rare time-domain events in large data sets and the development of fast ways to compare simulations quantitatively against data.

The ISM, star formation, and planet formation also involve key physical processes that are less commonly encountered elsewhere, including nonideal MHD and multiple-fluid effects, complex time-dependent chemistry and dust evolution, and planetary growth processes (F-Q1, F-Q2, F-Q3, F-Q4). Many of these physical processes have important effects on fluid motions, temperature, ionization, chemistry, and dust properties that are not yet adequately understood. Hundreds of diffuse interstellar bands are observed, but only a single carrier (C_{60}^{+}) has been identified, accounting for only three of the observed features (F-Q1). Uncertainties in collisional excitation cross sections limit the accuracy of temperature and abundance determinations from emission lines. Laboratory and atomic and molecular astrophysics needs to be supported to exploit and interpret astronomical data (F-Q1, F-Q2, F-Q3).

BOX F.1
Science Questions and Discovery Area: Interstellar Medium and Star and Planet Formation

F-Q1: How do star-forming structures arise from, and interact with, the diffuse interstellar medium?	F-Q1a: What sets the density, temperature, and magnetic structure of the diffuse ISM, enabling the formation of molecular clouds? F-Q1b: How do molecular clouds form from, and interact with, their environment? F-Q1c: How does injection of energy, momentum, and metals from stars (“stellar feedback”) drive the circulation of matter between phases of the ISM and CGM?

F-Q2: What regulates the structure and motions within molecular clouds?	F-Q2a: What processes are responsible for the observed velocity fields in molecular clouds? F-Q2b: What is the origin and prevalence of high-density structures in molecular clouds, and what role do they play in star formation? F-Q2c: What generates the observed chemical complexity of molecular gas?
F-Q3: How does gas flow from parsec scales down to protostars and their disks?	F-Q3a: How do dense molecular cloud cores collapse to form protostars and their disks? F-Q3b: How do protostars accrete from envelopes and disks, and what does this imply for protoplanetary disk transport and structure? F-Q3c: Is the stellar mass function universal?
F-Q4: Is planet formation fast or slow?	F-Q4a: What are the origins and demographics of disk substructures? F-Q4b: What is the range of physical environments available for planet formation? F-Q4c: How do turbulence and winds influence the evolution of structure in disks?
F-DA: Detecting and characterizing forming planets	F-DA1: How do planets and their satellites grow? F-DA2: What are the atmospheres of long-period giant planets like at their formation epoch? F-DA3: How do the orbital architectures of planetary systems evolve?

TABLE F.1 Summary

Capability	**Science Enabled**	**Current/Expected Facilities**	**Future Needs**
X-ray spectroscopy	F-Q1	Chandra, XMM	Observations of X-ray absorption fine structure from elements/minerals in dust and gas using 0.2–2 keV spectroscopy with high resolution (R ~ 3000) and 10–100× larger effective collecting area than current facilities.
UV spectroscopy	F-Q1, F-DA	HST	R > 10^5 for absorption lines of H_2, CO, and depletions (F-Q1); low R spectra to probe accretion from circumplanetary disks (F-DA)
Optical spectroscopy	F-Q1, F-Q3	Keck, Gemini, Magellan, MMT, Subaru, LBT	R > 6 × 10^4 spectra on ~8 m telescopes to map ISM absorption (F-Q1); R > 2 × 10^4 spectra on 8 m telescopes; rapid follow-up of transients (F-Q3)
High angular resolution optical-NIR imaging and	F-Q3, F-Q4, F-DA	VLT, Gemini, Keck, Magellan, MMT,	R ~ 4000 spectroscopy + imaging on ELTs (F-Q3); very high resolution

spectroscopy		Subaru, LBT	(coronagraphic) imaging of disks (F-Q4); very high angular resolution, low spectral resolution IFU spectra on ELTs (F-DA)
Near-IR spectroscopy	F-Q3, F-DA	VLT, Gemini, Keck, Magellan, Subaru, MMT, LBT	R > 2 × 10^4 spectra on 8 m telescopes (F-Q3); R ~ 100–1000 spectra on ELTs (F-DA)
Mid-IR imaging	F-Q3, F-DA	VLT, Gemini, Keck, Subaru	8 m and ELT imaging at ~ 3–20 microns (F-Q3, F-DA)
Far-IR spectroscopy	F-Q1, F-Q2	SOFIA	~1' polarized far-IR emission from diffuse ISM (F-Q1); 10" maps of polarized far-IR emission from filaments (F-Q2)
Sub/mm interferometry	F-Q2, F-Q3, F-Q4, F-DA	SMA, NOEMA, ALMA (JVLA)	Extragalactic < 1" maps of dense gas tracers, CN Zeeman in MW (F-Q2); very high spectral and spatial resolution of cores (F-Q3); very high spectral (tens of m/s) and spatial (milli-arcsecond) resolution measurements of molecular gas and dust in circumstellar and circumplanetary disks (F-Q4, F-DA)
Cm-wave single dish and interferometry	F-Q1, F-Q2, F-Q3, F-Q4, F-DA	JVLA, GBT, Arecibo	Deep (> JVLA) 21 cm 1" resolution imaging in MW and nearby galaxies (F-Q1); dense gas tracers (e.g., N_2H^+, NH_3) at <0.1 pc in MW; deep-line surveys (F-Q2); very high spectral and milli-arcsecond spatial resolution of cores (F-Q3); very high spectral (tens of m/s or better) and spatial (milli-arcsecond) resolution measurements of molecular gas and dust in circumstellar and circumplanetary disks (F-Q4, F-DA)
MHD+radiation hydro simulations, algorithm development	F-Q1, F-Q2, F-Q3, F-Q4, F-DA	XSEDE, NASA and DOE HPC	Multiphase galaxy simulations with feedback, CGM/ISM, cosmic rays (F-Q1); simulations of cloud to core scales with chemistry, nonideal MHD and radiation (F-Q2, F-Q3); simulations of disk-planet interactions, MHD turbulence, winds, circumplanetary disk evolution, planet accretion (F-Q4, F-DA)
Big Data	F-Q1, F-Q3	Data from Gaia, TESS, VRO, ZTF, etc.	3D ISM structure in gas, dust, and B field with Gaia (F-Q1); identification of accretion transients for rapid spectroscopic follow-up of VRO, ZTF, etc. (F-Q3)
Laboratory and other	F-Q1, F-Q2,	Existing laboratories	Properties of dust at X-ray through mm

theoretical studies: atomic and molecular spectroscopy, cross sections, physical properties of grain materials, plasma astrophysics	F-Q4	wavelengths (F-Q1); ices and gas/surface chemistry (F-Q2); ices, gas/surface chemistry, dust properties and growth (F-Q1, F-Q2, F-Q4)

G

Report of the Panel on Stars, the Sun, and Stellar Populations

INTRODUCTION

Stars are the main source of light in the universe. The scope of stellar astrophysics ranges from our host star, the Sun, to stars in galaxies so far away that we cannot distinguish them individually. Stars also span an immense range in mass and size. The least massive stars are barely distinguishable from giant planets, and have lifetimes that exceed the age of the universe. The most massive stars have lifetimes shorter than the history of humans on Earth, and die in spectacular explosions that seed the universe with heavy elements. Stellar astrophysics encompasses the study of single stars, stars in binary and multiple systems, and stars in bound clusters and associations that share common ages and compositions, a laboratory for investigating stellar evolution. Simple mass and energy conservation laws, nuclear fusion to generate energy, radiation and convection processes that transport that energy to the surface, and basic radiative transfer relations are sufficient to explain the most basic features of stars. Accurate modeling of these processes has allowed the study of stars to flourish, and has facilitated a transition from the mere cataloguing of observable phenomena to detailed study of fundamental physics. However, stars are not just static, spherically symmetric objects that are solely described by one-dimensional models; they exhibit spatially and temporally complex phenomena, from rotation and magnetic field generation to pulsations and explosions. These phenomena determine how the stars populate and evolve across the Hertzsprung-Russell (HR) diagram. Thus, while stellar astrophysics is an old endeavor, new observations and advanced theory and simulations continue to reveal new phenomena that push our understanding of the fundamental physical processes that generate all of the visible light of the universe.

The scope of the Panel on Stars, the Sun, and Stellar Populations includes stellar structure and evolution, stellar activity and variability, brown dwarfs, ground-based solar astrophysics, resolved stellar populations including star clusters, nucleosynthesis and chemical evolution. In the course of its work, the panel reviewed and incorporated the input of over 150 white papers submitted by the astronomy and astrophysics community addressing the preceding topics, as well as broader areas of astronomy in which stars are tools for studying fundamental physics, exploring the interstellar and intergalactic media, and probing distant galaxies. It is notable that most white papers had more than 10 authors and many were the result of organized efforts by sections of the community. The panel had a rich source of ideas and analyses to draw upon while developing its understanding of the needs and opportunities for advancing the science within its purview.

THE STATE OF THE FIELD

Before identifying the most critical problems in stellar astrophysics in the coming decade, the context is set by giving a brief overview of some of the most exciting discoveries in the past decade. This decade saw a resurgence of stellar astrophysics research fueled by time-domain observations by NASA's Kepler space telescope and ground-based networks, large-scale spectroscopic surveys such as Apache

Point Observatory Galactic Evolution Experiment (APOGEE) and Galactic Archaeology with HERMES (GALAH), sensitive and high-resolution sub-millimeter observations with Atacama Large Millimeter/Submillimeter Array (ALMA), and the unparalleled astrometric precision of Gaia. It truly was the "decade of stars." The study of the compositions, motions, ages, and multiplicity of stars has expanded from relatively small ($\approx 10^3$–10^5) to relatively large ($\approx 10^7$–10^9) samples exploring a large fraction of the Milky Way and its satellites, reaching to the brightest stars in galaxies throughout the Local Group and beyond. Our knowledge of the positions and motions of stars within the Milky Way has expanded ~1000-fold in number and ~10,000-fold in volume. The 2010 decadal survey[1] noted that time-domain astronomy would be the new frontier for stellar astrophysics. Kepler has opened that frontier, invigorating the field of asteroseismology that provides a window into the hidden interiors and fundamental properties of stars across the HR diagram, and revealing new categories of variable phenomena related to stellar evolution, magnetospheres, multiples, and circumstellar environments.

We start with the Sun. The goal of modern solar physics research is to understand the entire Sun, from the core to the heliopause, in order to provide a holistic description of variations in its magnetic fields and the associated eruptive phenomena that can affect life on Earth. Observations during the past decade have revealed that magnetic fields are the source of energy for solar flares, the heating of the solar corona, and the acceleration of solar wind particles that shapes the entire heliosphere. Chromospheric data from ALMA have dramatically improved our understanding of this key interface layer between the photosphere and the corona (see Figure G.1). Helioseismic observations, data that allow us to "see" inside the Sun, provided the path for solving the solar neutrino problem in the 1990s. These data are now used to study changes inside the Sun, and have revealed the beginning of a new solar cycle years before its first sunspots appear on the Sun's surface. Advances in the next decade should provide more detailed understanding about how solar magnetic fields drive energetic phenomena at different spatial, temporal, and energy scales.

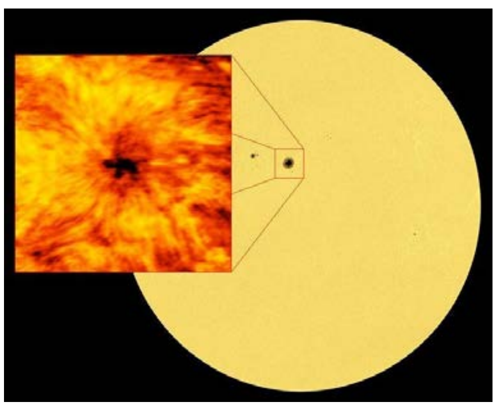


FIGURE G.1 An image of the entire Sun taken at a wavelength of 617.3 nm on December 18, 2015, showing light from the visible solar surface, the photosphere. The leading sunspot of AR 12470 is clearly visible on the disk. The inset is the same sunspot as seen in the chromosphere using ALMA observations at a wavelength of 1.25 mm. SOURCE: ALMA (ESO/NAOJ/NRAO); B. Saxton (NRAO/AUI/NSF).

The past decade marked a revolution in stellar astrophysics through large-scale asteroseismology. Data from NASA's Kepler satellite allowed, for the first time, systematic measurements of the radii and masses of thousands of individual stars, as well as age estimates. NASA's Transiting Exoplanet Survey Satellite (TESS) mission is now continuing this process through asteroseismic measurements of bright stars across the entire sky. Asteroseismology's greatest benefit is the access it provides to the internal structure and rotation of stars. The effect of internal rotation on the pulsation modes of evolved stars has allowed us to determine their core rotation rates precisely, and has revealed that we do not fully understand angular momentum transport during stellar evolution. These data have shown that angular

[1] National Academies of Sciences, Engineering, and Medicine, 2016, *New Worlds, New Horizons: A Midterm Assessment*, The National Academies Press, Washington, D.C.

momentum transport from core to surface is more effective than current theories had predicted, and have spurred the development of new interior evolution models. The ages provided by asteroseismic data have allowed us to determine stellar ages quite precisely, which can be used to quantify the age-rotation rate relation ("gyrochronology"), particularly for older stars. Again, these data have indicated that modifications are needed for current stellar spin-down theories, to explain the observed reduction in the efficiency of magnetic braking as a star ages.

Understanding of the connection between a star's magnetic field and its stellar wind is of profound interest for understanding a star's influence on its surroundings—that is, its astrosphere. Magnetic fields also affect a star's sphere of influence. Consequently, over the past decade, the concept of the habitable zone (HZ), where water could be in liquid form on an orbiting planet's surface, has now expanded to encompass the magnetic nature of stars and their influence on planetary evolution and the development of life. Understanding stellar activity and its high energy emission, both as a risk to, and as a possible catalyst for life, is now a central concept in astrobiology. While we have yet to measure a stellar wind for a cool star (although upper limits have been obtained for Proxima Centauri), and we have as yet only a few plausible claims of coronal mass ejection (CME) observations from stars other than the Sun, stellar flares have been observed frequently by Kepler, TESS, and ground-based monitoring networks.

The past decade has seen improvement in techniques that allow us to directly measure the fundamental properties of stars—mass, radius, and luminosity—even at the extremities of the mass spectrum. Such measurements for stars with masses too low to ignite hydrogen $\left(\approx 0.072\ M_{\odot}\right)$ have given us an empirical probe and new theoretical advancements on the equation of state of hydrogen-rich, electron-degenerate matter, and the role of magnetic fields in stellar structure. Indeed, the mass-radius relationship for the lowest-mass stars cannot be explained without taking both effects into account. The discovery and census of Y-type brown dwarfs with temperatures as cool as 250 K has produced the first measurement of the field "stellar" mass function down to planetary masses ($0.013\ M_{\odot}$).

At the other end of the mass spectrum, there has been significant progress in our understanding of how high-mass stars evolve. Binary interaction has been shown to be a common and critical ingredient in the evolution of massive stars. Surveys show that interacting binaries are ubiquitous among these stars, accounting for at least 70 percent of the population, and that mass exchange can substantially modify their evolutionary trajectories. Furthermore, as massive stars experience substantial mass loss both on and after the main sequence, the uniformity and regularity of mass loss can significantly modify evolutionary time scales. Extreme or episodic changes in loss rates can significantly hasten the evolutionary time scale to a star's end phases. We have also learned that about 10 percent of massive stars possess strong surface magnetic fields (~0.3–20 kG) and high rotation rates (> 200 km/s at the equator); both factors have a strong influence on mass loss and accretion. Stellar population synthesis models have improved by accounting for these factors, and has led to a better accounting of the ionizing output from massive stars. Stellar evolution models are also better at quantifying the connection between the physical properties and multiplicity of massive star systems and the types of supernovae they produce, as well as the types and masses of compact remnants they leave behind. The latter are particularly relevant for understanding the rate and nature of compact-object mergers that are now regularly being detected by Laser Interferometer Gravitational-Wave Observatory (LIGO).

Time-domain surveys have opened up the study of the end states of stars through the detection of explosive, transient phenomena. All-sky surveys with nightly cadences have revealed several new classes of transients, including high-energy but short-duration (<1 day) relativistic explosions, long-duration but intermediate-luminosity red transients, and a myriad of faint thermonuclear supernovae (SNe) that range from calcium-rich transients to subluminous Type Ia and Iax SNe. These discoveries are not only enhancing our understanding of the variety of ways stars die by filling in the empirical picture of the end stages of stellar evolution; they are also probing detailed stellar processes such as mass loss, binary evolution, and interactions with the interstellar medium (ISM).

The deployment of moderate-size spectroscopic surveys measuring stellar heavy-element abundances and velocities, combined with stellar astrometry from Gaia and ages derived from asteroseismic data, has extended our acuity of stellar populations beyond the solar neighborhood,

resulting in major revisions in our understanding of the formation and chemical evolution of the Milky Way. Theories of galaxy formation suggest that galaxies such as the Milky Way grew by accreting gas and stars, and in the past decade we have discovered key facts about how matter accumulates and gas turns into stars to create the visible Milky Way. Data from Gaia have allowed the identification of a large, ancient accretion event, consisting of stars that dominate the inner stellar halo of the Milky Way. Ages of stars present in the Milky Way disk show that the disk was formed "inside-out," with star formation starting earlier in the inner regions where gas first accumulated, and thus the Milky Way's disk reached its current size gradually. Comparable photometric and spectroscopic studies of individual stars in local group dwarf galaxies are enhancing our knowledge of the origins and evolution of these satellite stellar populations.

The data on chemical abundances have shown that the fraction of stars whose chemical enrichment was dominated by elements formed in massive-star supernovae, compared with those that show enrichment by both massive-star and white dwarf supernovae, varies enormously across the disk of the Milky Way. However, these groups cannot be separated clearly by age or overall heavy-element enrichment. There is no consensus yet on the cause of the abundance variations, but hypotheses include the radial movement of stars from the inner part of the Milky Way outward, and a burst of massive star formation following the infall of fresh gas into the Milky Way.

Stellar astrophysics research over the past decade has also been revolutionized by large open-access data sets from numerous photometric, astrometric, and spectroscopic surveys. While many publicly funded heliophysics and space-based facilities have had open-data policies for many years, such policies are reasonably new for ground-based astronomy. The exploitation of the new extensive photometric and spectroscopic data sets has been greatly aided by new computational methods, such as machine learning and data-driven modeling, that are now regularly employed, as well as by publicly available software such as Modules for Experiments in Stellar Astrophysics (MESA) and Astropy. Additionally, improvements in molecular and atomic line-lists, and continued work on 3D and nonequilibrium effects in model stellar atmospheres, have been important in the past decade. Nevertheless, inputs to stellar models are still quite uncertain, particularly opacities of molecules and highly ionized atomic species. Opacities used in stellar models are often produced through computationally intensive calculations, which in some cases are difficult to calibrate and test. For molecules, the constraints are often numerical, with line lists exceeding 10^{10} transitions. For ionized atoms, the constraints are often experimental, requiring extreme testing conditions. For example, experiments at Sandia National Laboratory have revealed discrepancies between the measured and calculated opacities of iron, nickel, and chromium at conditions found at the base of the solar convection zone. The position of the solar convection-zone base is known precisely ($0.713 \pm 0.001\ R_\odot$) from helioseismic analyses, but in solar models this depends on opacities and models created with the latest estimates of solar heavy-element abundances that have convection-zone bases that are discrepant at the 11σ level. While discrepancies between theory and experiment were anticipated, the reasons for these discrepancies are not yet understood.

While the past decade has witnessed many significant advances in the study of the Sun, stars, and stellar systems, many outstanding questions remain. These include questions about the fundamental mechanisms driving magnetic field generation and its influence on internal structure, surface heterogeneity, wind generation, and environmental influence across the mass spectrum; the temporally and spatially dynamic properties of stars, and their internal to external manifestations; and the formation and co-evolution of stellar multiples, and their influence on the end states of massive stars. Moreover, with new technologies, computational tools and theoretical developments, we have the opportunity to build a complex and comprehensive assessment of stellar populations throughout the Milky Way and beyond. The most promising scientific opportunities are described in the next sections.

THE SCIENCE PRIORITIES

The following sections describe what the panel believes are the science priorities of our field. The four science questions whose answers will bring transformative change to this field are (1) What are the most extreme stars and stellar populations? (2) How does multiplicity affect how a star lives and dies? (3) What would stars look like if we could view them like we do the Sun? (4) How do the Sun and other stars create space weather? The discovery area that will give the highest science return is “industrial scale” spectroscopy of 10^9 stars in the Milky Way and beyond.

G-Q1. WHAT ARE THE MOST EXTREME STARS AND STELLAR POPULATIONS?

Stars at the limits of mass, composition, pulsation properties, and rotation test the ability of theoretical models and also have profound implications for the evolution of galaxies. Stars that are not in equilibrium as they transition rapidly between more static states have rarely been seen, even though all stars go through such phases. Defining the extremities is also crucial for stellar systems because they can span a wide range in properties such as masses, rotation, magnetic fields, mass transfer rates, and so on. For star clusters and stellar populations, the range of the distribution of masses and frequency of multiple stars can be determined from observations and these properties play a critical role in the evolution of the system. Over the next decade, Hertzsprung-Russell (HR) diagrams of the Milky Way from Gaia and of the local universe with James Webb Space Telescope (JWST), combined with theoretical explorations, will revolutionize our understanding of stars and stellar populations.

More than a century since its creation, the HR diagram—a plot of luminosity against surface temperature—remains one of the most powerful tools for understanding stars and stellar systems. A star’s location on the HR diagram encodes essential physics including the star’s current energy source, internal structure, elemental abundances, and circumstellar environment. Patterns in the HR diagram reveal deep truths about how stars evolve, as well as the ages and chemical abundances of the environments that host the stars. Mapping and interpreting the deep connections between the HR diagram and stellar physics have impacts throughout astrophysics, from reionization in the early universe, to the history of galaxies, to the fundamental physics of degenerate matter. In order to answer the question above, we need to determine the fundamental properties and demographics of stars across the HR diagram, and what these properties imply for the evolution of stars.

The past decade has transformed our characterization of the HR diagram within the Milky Way. The astrometric Gaia mission measured distances to millions of individual stars, placing them securely on the HR diagram and revealing stars that are unquestionably in unexpected regions of the HR diagram. Cross-referencing the Gaia HR diagram to extensive photometric and spectrographic catalogs (many the products of the past decade’s investments) has made it possible to identify theoretically predicted but rare phases of stellar evolution, such as the aftermath of mass transfer between binary stars. It is in this mapping between the observable features of the HR diagram and the fundamental parameters of stars that the next decade will shine. We now know empirically how stars of different types populate the HR diagram, but the connection to physical parameters and the underlying stellar physics is often limited in precision, completeness, accuracy, dynamic range, and theoretical understanding. The lack of precise stellar masses and radii remains a limiting factor for exoplanet characterization, for studies of the equation of state for degenerate matter, and for understanding the stellar variables that anchor much of the extragalactic distance ladder. For example, to study the influence of convection on stars, and in particular how to implement it in models, requires radius estimates to better than 5 percent. Understanding the inflated radii of M dwarfs requires even better precisions, of the order of a few percent. At the low-mass end, mass estimates to better than 10 percent are needed to understand the transition between brown dwarfs and stars. Lack of accurate masses, binarity, and rotation limits our knowledge of stars that explode as supernovae or erupt as other violent astrophysical transients. Incomplete stellar samples limit our knowledge of the initial mass function, and may miss rapid evolutionary stages that dominate a stellar

population's overall luminosity, supernova rate, and dust production. Precise physical parameters are also needed to calibrate atmospheric and evolutionary models that are used to make age determinations and to trace chemical enrichment of stellar populations. All of these issues are least understood at the extremes of stellar mass and chemical composition, preventing a comprehensive theoretical understanding of stellar interiors, atmospheres, nucleosynthesis, and evolution.

The path forward lies in harnessing both old and new approaches on a large scale. Fundamental measurements of mass and radius can be made directly, but are typically limited to relatively bright stars for which asteroseismic or interferometric data can be obtained, or to rare varieties of stars, such as eclipsing binaries. Expanding this to statistically powerful samples across the HR diagram can be done by expanding the eclipsing and noneclipsing binary sample and extending direct interferometric radius measurements to single stars at low masses and beyond the solar neighborhood. Continuing a vigorous program of asteroseismology, which has revolutionized understanding of solar-type and more massive stars, will also extend precision calibrations to lower masses and additional stellar types. These results can be used to calibrate indirect methods to infer mass and radius (i.e., theoretical models). Microlensing mass measurements with Wide-Field Infrared Survey Telescope (WFIRST) will extend calibrations across the full stellar mass spectrum, provided that lensing stars can be detected and photometrically anchored to the HR diagram. Laser Interferometer Space Antenna (LISA) could help in determining the radii of the extreme low-mass end of stars.

Interpreting the HR diagram also requires accounting for the many factors beyond mass that influence stellar luminosities and atmospheric temperature. These factors may include age, chemical composition, nonequilibrium atmospheres, rotation, binarity, and so on. Direct age measurements can come from nucleo-cosmochronology for old metal-poor stars, lithium depletion for young fully convective stars and brown dwarfs, asteroseismology, and age-sensitive abundance indicators. These age standards can calibrate gyrochronology, and HR-based age-dating methods can anchor theoretical stellar evolution models. Accurate interpretation of the HR diagram requires stellar samples spanning a wide range of ages and elemental abundances, spectral data outside the optical band, better atomic and molecular opacities, improved nonequilibrium atmosphere modeling and retrieval methods, and an accounting for evolutionary effects (e.g., self-enrichment).

Rotation and multiplicity remain significant sources of uncertainty in interpreting the HR diagram, and there is a pressing need to determine stellar rotation rates with synoptic time-domain observations, asteroseismology, and spectroscopy. Such observations will help improve models of how stellar rotation evolves, as well as models of binary co-evolution. The other unknown is the magnetic field. While proxies of chromospheric magnetic activity have been monitored for decades, measurements of magnetic fields and their configurations remain rare. These measurements are becoming more accessible through spectropolarimetry and Zeeman Doppler Imaging (ZDI), but the resources remain scarce.

Expanding the depth of our knowledge of the physics that underpins the HR diagram not only deepens our knowledge of stellar processes, it also unlocks the HR diagram's power for *all* of astrophysics. The HR diagram remains the most accessible constraint on stars' properties and evolutionary state, given the greater observational demands on spectroscopy or asteroseismology. Completing the mapping between the HR diagram and internal stellar physics leverages information that is only observationally available in the Milky Way or its satellites. This mapping will be critical as we resolve individual stars at mega-parsec distances with JWST and 30 m class telescopes, and study more distant stellar populations through their integrated light.

G-Q2. HOW DOES MULTIPLICITY AFFECT THE WAY A STAR LIVES AND DIES?

Most stars orbit other stars. Those in very widely separated systems are crucial probes of star formation processes and coeval laboratories for stellar structure and evolution studies, and constrain the properties of the Milky Way's dark matter. Those in very closely separated systems have fates that are

intertwined, altering each other's evolution, mass loss, and final outcomes. Close binary interaction can manifest itself as transients associated with mass transfers and mergers. Binary co-evolution fundamentally influences nucleosynthesis, governs supernovae rates, affects gravitational wave production, and modifies stellar lifetimes. Binary co-evolution also influences the structure of planetary nebulae, the demographics of compact objects as well as our interpretation of the HR diagram. While the past decade has driven home the ubiquity of these phenomena, the demographics and mechanisms of binary co-evolution remain weakly constrained. Better observational samples and theoretical modeling of these phenomena are the route to addressing some of the more challenging questions such as the fate of close binaries that do interact.

Observations to date show that over 70 percent of stars above 8 $M_{\odot}$ are in multiple systems. Being bright, these can outshine other stars; such systems are believed to dominate the UV light from external galaxies. High-mass stars in binaries can interact before becoming supernovae, as the η Car system demonstrates. Many multiple systems have at least some mass transfer or are in close contact with one another, while others co-evolve through a poorly understood common envelope configuration, or even merge. There can be substantial changes to the emergent radiation of a star if its outer layers are stripped off by a companion. Indeed, binary interactions in very metal-poor stars may provide enough high-energy photons to contribute substantially to the reionization of the universe. Stripped envelopes and other mass-loss debris surround a dying star and can interact with supernova ejecta. Most of the diversity in core-collapse supernova light-curves may be owing to interactions with circumstellar material rather than the underlying engine; to understand the physics of the latter, we must understand the phenomenology of the former. For these reasons, it is imperative that we correctly determine the fraction of high-mass stars in binaries, particularly close binaries for which co-evolution effects are most prominent; and measure the outcomes of binary co-evolution, such as mass-transfer rates.

Stars with masses less than 8 $M_{\odot}$ can also have close companions, creating the progenitors of cataclysmic variables and Type Ia supernovae. The progenitors of these transients co-evolve in at least one common envelope phase. Depending on the initial masses, separations, and system mass-loss rates, they evolve into either a single white dwarf star with mass accretion from an evolved main sequence or red giant companion, or into a pair of white dwarf stars. As with massive stars, the details of what occurs in the common-envelope phase are not well understood, but they are critical for determining whether the outcome is capable of creating a Type Ia SN, and whether we observe it earlier as an AM CVn star, a supersoft X-ray source, or a double white dwarf system. Stars between 0.5 and 1.5 $M_{\odot}$ are prime candidates for habitable worlds because of their long lifetimes and relatively benign astrospheres. Roughly two-thirds of Sun-like stars are in binaries, thus understanding how planet formation and planetary-system evolution is affected by the gravitational and radiation influences of other stars in a system is crucial. We know that planet formation in binary systems is possible, but it may be disfavored. Depending on the proximity of a stellar companion and the stellar mass ratio, the companion can influence the dynamical evolution of a planetary system through processes such as Kozai-Lidov oscillations. Understanding the role of multibody interactions in planetary systems requires identifying the frequency of such systems and the correlation of stellar binarity with the architecture and orbits (inclinations and eccentricities) of planets.

Despite the prevalence and importance of multiple-star systems, we have not yet mapped their demographics sufficiently. Among other parameters, we need to know the orbital properties and mass ratios of multiple-star systems across age, mass, mass-loss rates, and composition. While we know that multiplicity is more common in high-mass than in low-mass stars, we have incomplete information on the statistics of systems with extreme mass ratios, long orbit periods, and very low masses (e.g., brown dwarfs). The first investigations of how multiplicity depends on composition have indicated that low-mass metal-poor stars have higher rates of multiplicity. Further investigation of this population is necessary because multiplicity at very low metallicities is relevant to the ionizing flux, nucleosynthesis, carbon abundances, and pollution signatures of Pop III stars. Measurement of multiplicity in systems losing large amounts of mass is needed to forecast mass-loss rates for broader stellar populations.

Information on systems with more than two stars is also lacking, despite their potential importance in Type Ia production and star-planet dynamics. With current and upcoming survey capabilities, we are poised to map multiplicity in exciting new ways, using astrometry from Gaia, gravitational waves from LIGO and LISA, and synoptic photometry from the Zwicky Transient Facility (ZTF), Legacy Survey of Space and Time (LSST), TESS, and others. The addition of spectroscopic capabilities is crucial to the prospects for discovery and characterization.

Efforts in theory and computation need to go hand in hand with observations. The vast majority of stellar structure and evolution calculations assume single stars. There are very few modeling efforts that take binarity into account, and fewer still that take both rotation and binarity into account. Given that such an effort will need multidimensional models of multiple, interacting stars, there needs to be support for new code-development in this field.

G-Q3. WHAT WOULD STARS LOOK LIKE IF WE COULD VIEW THEM LIKE WE DO THE SUN?

With the exception of Betelgeuse, a supergiant that has a radius almost 1000 times that of the Sun, we cannot easily observe features on a star's surface. As a result, stars are typically treated observationally as featureless points of light, and theoretically as spherically symmetric objects. However, observations of our own Sun demonstrate that stellar surfaces are complex and dynamic, and sometimes not even spherical. Spots, flares, tidal distortions, mass-loss, rotation, and internal convection all break the spherical symmetry of stars. Observations in the past decade have revealed the pervasiveness of these effects, which have confounded our ability to understand the fundamental properties of stars and their surroundings. Advances in computing and computational methods are now beginning to make sophisticated 3D models of stellar interiors and atmospheres a possibility; this will enable greater clarity on a myriad of stellar asymmetries.

Physical processes that break interior symmetry, such as rotation and meridional flows, cause mixing across chemically inhomogeneous interior layers, and this in turn alters how these stars evolve. These effects can only be approximated in currently used 1D models, often leading to conflicts with precise observations. Convective heat transport is another source of error in models. Helioseismic and asteroseismic data have already shown that 1D approximations of convection do not model the surface layers of stars correctly. More worryingly, free parameters in 1D approximations of convection and magnetic field generation can lead to incorrect predictions of stellar radii. Helioseismic data also reveal extra mixing below the solar convection zone, a feature not present in standard 1D models. Asymmetric processes affect the later evolutionary stages as well. For instance, the internal mass distribution of white dwarfs, which informs models of supernovae and chemical enrichment, depends on the sizes of the cores in their progenitors. However, core size is affected by convective overshoot, rotationally induced mixing, and related instabilities that are inherently 3D in nature, and cannot be understood with a 1D approach. 3D models are also needed to properly understand the interplay between stellar convection, rotation, and magnetic field generation.

At the stellar surface, where photons can escape and reach us unimpeded, stellar asymmetry manifests itself in the form of spots: large areas of enhanced magnetic field and generally reduced brightness, with enhanced brightness in the surrounding "plage." These dark spots were the features Galileo identified as blemishes on the Sun. In the past decade, we have witnessed how varied the nature of star spots can be. Even stars with masses similar to our Sun can have dramatic differences in spot sizes and surface distributions. For lower-mass stars, spots often occupy a much larger fraction of the stellar surface and are not limited to low latitudes. Over the next decade, it will be critical to develop a comprehensive understanding of how star-spot sizes and distributions are driven by the underlying magnetic structure, and physical properties, of stars of all types. In terms of theory, complete 3D magnetohydrodynamic models of stellar dynamos are essential to understand the details of stellar magnetic configurations and their time variation, particularly in fully convective stars.

The Sun remains our best model in understanding surface inhomogeneities, and new data from the Daniel K. Inouye Solar Telescope (DKIST) will revolutionize the field by allowing us to infer details of magnetic field strengths and orientations, as well as magnetic processes such as reconnection, in the solar atmosphere. These discoveries need to be expanded to encompass other stars through high-sensitivity stellar spectropolarimetry, which measures a star's magnetic field structure and plasma properties. Because spot asymmetries can masquerade as exoplanet signals, velocity-resolved stellar surface spectra (Doppler imaging) will provide a critical reference for disentangling exoplanet and stellar signatures, essential for the robust detection of habitable Earth-like planets.

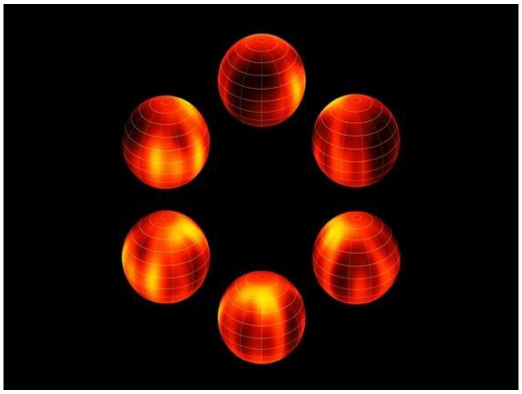


FIGURE G.2 A weather map of WISE J104915.57-531906.1B, informally known as Luhman 16B. This is the nearest brown dwarf to Earth and was discovered using NASA's WISE mission. ESO's Very Large Telescope was used to create this map of the weather on the surface of this brown dwarf. The figure shows the object at six equally spaced times as it rotates once on its axis. SOURCE: ESO/I. Crossfield.

In the cool atmospheres of brown dwarfs, where magnetic spot formation can be inhibited, surface asymmetries arise instead from condensate cloud structures (see Figure G.2). The complexity of the observed light curves of cool brown dwarfs points to global dynamic processes similar to those observed in the solar system giant planets, which may include thermochemical instabilities that can trigger (or mimic) cloud structure. Better understanding of brown dwarf cloud structure and composition through spectroscopic monitoring and 3D circulation and condensation modeling will also contribute to our understanding of exoplanet atmospheres, which have also shown evidence of cloud-induced variability.

Away from the stellar surface, asymmetries are conspicuous as nonisotropic, episodic, and clumpy mass outflows that emerge from stars of all masses. Low-mass stars generally lose large amounts of mass only during their late evolutionary phases on the asymptotic giant branch (AGB), whereas high-mass stars lose mass throughout their entire lives. Among the challenges we face in interpreting stellar outflows are those of understanding the rate of radiation-driven mass loss in high-mass stars, understanding the wind-launching and dust acceleration mechanisms in AGB stars, and deriving robust empirical estimates for mass loss in the late stages of massive star evolution (luminous blue variables, red supergiants). These observational effects need to be matched to reliable theoretical predictions. Because mass loss has a deterministic influence on the evolution of evolved stars, it is increasingly important to understand the nature of clumpy, episodic, and sometimes eruptive states of mass loss. These are the far more dominant modes of mass loss than the weaker winds typically associated with mass loss in solitary luminous stars. Current models have difficulty in predicting mass loss. There is a need for the development of quantitative theoretical predictions of mass loss rates in red supergiants and the underlying physics driving them. This is one of the main issues in understanding the evolution of massive red supergiants like Betelgeuse.

Understanding the evolution of various astrophysical phenomena, from exoplanet atmospheres to solar-like magnetospheres to the interaction of core-collapse SNe with circumstellar medium requires an

appreciation of the inherent asymmetries present from the interiors to the surfaces to the winds of stars. These require study in a comprehensive 3D manner.

G-Q4. HOW DO THE SUN AND OTHER STARS CREATE SPACE WEATHER?

The Sun and other stars affect their environments in numerous ways, from the interaction of stellar winds, flares, coronal mass ejections (CMEs) and other forms of mass loss with surrounding disks, planetary bodies, stellar companions, and the interstellar medium, to the creation of planetary nebulae and supernova remnants and the end stages of stars. We broadly interpret these phenomena as space weather, expanding on the traditional definition of this term to incorporate a star's influence on its environment throughout its life cycle. Understanding how a star generates these phenomena helps explain how stellar environments evolve over the star's lifetime and beyond. Identifying the physical processes involved in these interactions informs a broad range of current astrophysical problems, from stellar feedback in galaxy evolution, to the formation and retention of atmospheres on planets. Now is the time to address this question as new capabilities allow us to observe embedded hot stars in star-forming clouds, CMEs beyond the Sun, and the diversity of exoplanets orbiting other stars.

The physical scale of a star's influence on its environment is a function of its mass, with both quiescent and transient effects playing significant roles. Stellar radiation across the electromagnetic spectrum can have profound effects on the evolution of structure—for example, the photoionization of star-forming regions by the UV emission of a single massive star. Stellar mass loss through steady, episodic, or transient processes also influences the star's environment. The radiation-driven winds of massive stars interact with the circumstellar environment, creating nebulae filled with gas and dust. A massive star's evolution depends on its mass loss rate, which in turn depends on metallicity as the wind acts on the highly ionized metals produced by the star, and on magnetic field strength, now measured in 10 percent of hot stars. The winds of low-mass stars are more elusive, but the solar system provides an essential laboratory for understanding these processes. We know that the magnetically driven solar wind deflects the tails of Sun-grazing comets, drives atmospheric mass loss, and interacts with planetary magnetospheres, generating aurorae. The multiphase solar wind must have an acceleration mechanism beyond thermal expansion. The steady, fast solar wind originates from open magnetic field lines near the polar regions; the origins of the slower, more variable solar wind are still unclear, but seem to be related to the opening of field lines associated with active regions and coronal loops. Synoptic surveys are essential for modeling these processes. Solar flares probe fundamental particle acceleration, although the specific nature of magnetic acceleration requires further advances in both modeling and radio polarimetry. CMEs, containing lower-temperature, higher-density material, and generating the most damaging aspects of space weather for Earth, often accompany flares, although our understanding of the relationship is incomplete. Critically, all of these energetic phenomena originate from magnetic processes on the Sun, and we will soon be able to measure details of the magnetic phenomena powering these events with DKIST.

For low-mass stars, similar phenomena seem to occur, although often with very different properties. Stellar flares on active stars extending to the lowest stellar masses, observed from the radio to the X ray, can have high-energy luminosities up to five orders of magnitude larger than flares on the Sun. However, current observations tend to be biased toward the nearest and/or most active stars. In the next decade, we need to characterize transient energetic phenomena systematically, as a function of stellar type, mass, age, metallicity, and rotation rate, and to explore these phenomena across the electromagnetic spectrum. Direct observation of stellar winds and CMEs will help constrain these effects. Possible paths forward include CME detection through photoelectric absorption of a star's X-ray spectrum, and high-spatial/high-throughput X-ray imaging of the asteropause shock to measure wind kinematics.

Understanding the energetic phenomena driven by stellar magnetism requires exploration of the diversity of stellar magnetic surface structures. Magnetic field maps constructed using Zeeman Doppler Imaging (ZDI) show that nondipole fields can dominate the surface field structures of some stars, but

measurements have been made for only a few systems. The poles of the young main sequence star AB Dor appear to have the same polarity, based on contemporaneous Doppler imaging and X-ray spectroscopy and light curves. X-ray studies with high spectral resolution and throughput, combined with ZDI, are needed for a larger sample of low-mass stars at various evolutionary stages.

Intertwined with stellar magnetism is the long-standing question of what heats the stellar corona. There are two equally plausible mechanisms within steady-state magnetic loops: magnetic wave heating and flare-like impulsive heating scaled down to nanoflares. Distinguishing between these mechanisms requires systematic characterization of stellar coronal properties across a range of energies, as well as theoretical modeling that extends down to low-mass stars. Indeed, evidence from the few stars that could be observed spectroscopically with current X-ray technology suggests a diversity of coronal heating mechanisms, and as such the situation may be far more complex than currently appreciated. Very active stars show high electron densities in relatively quiescent coronae, densities observed only in flares on the Sun. Furthermore, the decline in optical/X-ray magnetic emission and its decoupling from persistent (and variable) radio emission at the lowest stellar and substellar masses suggests a profound change in magnetospheric structure and heating mechanisms. Coherent pulsar-like radio pulses and highly polarized emission suggest a connection with Jovian-like auroral mechanisms, but many mysteries remain. The detailed mechanisms of magnetic emission around low-mass stars have clear implications on the habitability of close-in orbiting exoplanets.

Ultimately, processes interior to the star generate the magnetic dynamos observed through their energetic surface phenomena. The growing interest in exoplanet atmospheres and potential habitability mandates a better theoretical understanding of stellar magnetism and its effects throughout a star’s system. Detailed magnetohydrodynamic dynamo models of the Sun and other stars spanning a range of physical properties are needed to explore these effects. These include investigation of potentially novel dynamo processes in fully convective stars. However, dynamo models are not enough. These need to be coupled to models of radiative transport through a star’s exosphere to predict emergent phenomena.

G-DA. DISCOVERY AREA: “INDUSTRIAL-SCALE” SPECTROSCOPY

Building on the scientific progress from large-scale, time-domain photometric surveys over the past decade, the panel sees that considerable advancement can be made by greatly expanding spectroscopic surveys in breadth, sensitivity, precision, and cadence across the full electromagnetic spectrum, or in other words extremely large-scale or “industrial-scale” spectroscopy. This capacity will be accomplished through advancements and investment in instrumentation and facilities, improvements and standardization of spectroscopic reduction and analysis techniques, archival, and broad community access to data products, development of novel approaches to explore the highly multidimensional data sets that will emerge from these efforts, and support for laboratory astrophysics, including theory and experiment.

Astronomy became astrophysics with the first spectrum. Spectroscopy determines compositions, magnetic field strength, space motion, rotation, multiplicity, planetary companions, surface structure, and other important physical traits. Industrial-scale spectroscopy expands current capabilities in spatial, spectral, temporal, and sample-size dimensions, with a higher sensitivity that enables deeper, farther, and faster observations. Spectroscopy is too important to continue to be a “follow-up” of photometric surveys. In the next decade, spectroscopy will be the dominant discovery tool for astronomy. The need for photometry will not go away; we will still need Kepler-like stable time-domain photometry that is combined with the sky coverage of TESS; we also need SPHEREx-like broadband spectral fluxes.

In the X-ray regime, new advances will come from expanding the rate and wavelength coverage of spectroscopic observations. Increases in throughput by two or more orders of magnitude, and increases in spectral resolution by factors of 2 to 5, will open up discovery space through measurements of thermal broadening and new line-ratio diagnostics. Athena does not have all the characteristics needed for work on stars. While Athena’s microcalorimeter will provide high-resolution spectra for the most energetic phenomena such as stellar flares, studies of quiescent and/or lower energy phenomena, such as coronal

heating and absorption of CMEs, would benefit from high spectral resolution at lower energies; these can be obtained with grating spectrometers. Furthermore, Athena does not have sufficient spatial resolution to disentangle individual stars in star-forming regions or other crowded fields. The current sample of stars with X-ray grating observations is less than 200, so an increase in sample size by two orders of magnitudes or more will dramatically change the field. For low-mass stars, high spectral resolution for dozens of M dwarfs will probe the magnetic transition to fully convective stars. For high-mass stars, time and spectrally resolved X-ray line profiles will probe the clumpiness of shocks in the wind and constrain mass loss rates.

Past UV spectroscopic samples are small, both in number and in wavelength coverage. Extending our reach to the extreme UV opens up discovery space in stellar exopheres, magnetism, and CME detection on stars other than the Sun by allowing the study of lines such as FeIX, FeX, FeXI, and so on. Providing multiplexing capability in the near UV is critical for observing those few stars in large samples that trace nucleosynthetic enrichment from the first generation of stars, as well as from merging neutron stars.

Moderate increases in the number of optical and near-IR stellar spectra provided by Milky Way surveys on 2–3 m class telescopes, and by Local Group surveys on 10 m class telescopes, have already changed the fields of stellar physics, Milky Way archaeology, and the origin of the elements. Further breakthroughs in these areas will be fueled by (1) very large samples—10^8–10^9 stars—on current telescopes and (2) wide-area surveys on even larger telescopes. These observations will yield results on extremely faint stars, as well as on extremely rare (e.g., hyper-metal poor) stars. In combination with Gaia astrometry, we will more fully understand how chemical evolution works on galaxy-wide scales in the Milky Way. Non-U.S. projects such as WEAVE and 4MOST will be a step in this direction, but neither will reach the scale that can transform the field completely; an order of order of magnitude increase in distance sensitivity would permit robust statistical studies of populations that are currently undersampled and will also probe flare energetics to the point where they merge with quiescent-scale emission and solar-like flares.

Much progress can be made using broadband radio measurements. Radio wavelengths can be used as a probe of particle acceleration. A continuous frequency coverage from 10–400 GHz on sufficiently large baselines can measure mass-loss rates in evolved stars. For the Sun, increasing observational cadence without sacrificing signal to noise will enable us to measure spectral changes quickly enough to study solar flares, CMEs, and shocks.

We know the expected outcomes of industrial-scale spectroscopy. However, just as extensive time-domain photometry revealed unknown categories of transients, transiting debris, unusual variables, and megaflares on solar-type and low-mass stars; we anticipate that changing the scale of spectroscopy will reveal new, unanticipated phenomena. Even when we know the general physics that can be probed with these observations, orders-of-magnitude improvement in sample size will provide new, unanticipated insights about the universe, as the gains made from asteroseismology have shown. Information across many bandpasses—for example, with SphereX in the mid-IR—will further enrich this discovery area. The full impact of a decade of spectroscopy-driven discoveries will require a multiwavelength approach to observations and modeling. Because there are few nationally supported facilities with any of the capabilities highlighted above, this discovery area will require investment to reach its potential.

Required Capabilities

Given the wide range of stellar temperatures, brightness, and environments, the potential advances identified in this appendix require a multifaceted approach encompassing observing, computing, and laboratory capabilities and resources.

Long-term global and synoptic monitoring of the Sun, in optical and radio wavelengths, coupled with detailed DKIST and space-based observations, is necessary for dynamic helioseismology and magnetographic monitoring of the full solar disk. The construction of the 4 m DKIST is almost complete

and will start collecting science data soon, providing us with detailed studies of magnetic interactions. However, observing with DKIST will be like watching the Sun through a microscope. The telescope will provide observations of magnetic fields in exquisite detail, *but these observations need to be put into a global context*, and this will require new synoptic facilities that measure the global magnetic fields of the Sun at high cadence, as well as helioseismic observations that will allow us to determine associated changes in solar structure and dynamics below the photosphere. Radio facilities are needed to probe changes in the chromospheric magnetic field. Coronagraphs capable of measuring coronal magnetic fields are required to obtain a complete picture of the entire magnetic Sun; current coronagraphs measure only intensity.

As far as asteroseismology is concerned, the large pixel sizes, and short time series, of TESS means that the progress that was being made in characterizing the global properties and internal structure of stars has slowed down. One way to make progress is to have another space mission that can make high-cadence observations of stars down to $V > 16$ magnitude for more than 4 years, with pixel sizes small enough to ensure contamination-free observations of dense stellar systems like cores of star clusters.

Highly multiplexed, panchromatic, spectroscopic surveys are needed to obtain the precise temperatures, luminosities, elemental abundances, and velocities—and their variance over time—for the number and variety of stars to be investigated in the coming decade. Multiplexed instrumentation matches the scale of current photometric ($\approx 10^{11}$ sources) and astrometric ($\approx 10^{9}$ sources) samples, and enables the discovery of rare stellar classes and short evolutionary phases that push the limits of astrophysical theory. High-sensitivity (large aperture) spectroscopy extends stellar measurement far beyond the Milky Way, and enables study of intrinsically rare stars that are unlikely to be close. Except for the case of very faint stars, where ELTs are required, the key to obtaining such large numbers of spectra is new instruments that can be put on 4–10 m telescopes. Panchromatic spectroscopy, facilitated by advanced UV and X-ray facilities, probes the full range of stellar phenomena, including coronal heating and mass ejection processes in the X ray, mass loss from massive stars and atomic abundances of metal-poor stars in the UV, and the molecular chemistry of very cool and highly embedded stars in the infrared. Multi-epoch observations map stellar binaries and reveal invisible companions, probe stellar interiors through precision asteroseismic measurements, and unveil the dynamic atmospheres of cloudy brown dwarfs and massive evolving stars. Concurrent advances in spectropolarimetry are needed to map stellar magnetic field structures, particularly for the lowest and highest mass stars whose interior structures differ considerably from the Sun's and whose interior dynamos remain poorly understood.

High-resolution imaging and interferometry are needed to resolve individual stars and stellar systems at relevant scales: 10–100 milliarcsecond optical and infrared imaging will isolate stars below the main-sequence turnoff in the Local Group, resolve and map the orbits of tight binaries, and enable proper motion selection in distant clusters; 10–100 microarcsecond near-IR imaging will provide parallaxes of cool and deeply embedded stars at kpc scales. Sub-microarcsecond interferometry will directly measure stellar radii down to substellar masses, monitor structural distortion in evolved stars, and resolve massive multiples in distant clusters. High-resolution imaging and imaging spectroscopy, and spectropolarimetry in the optical and infrared, will create direct probes of thermal and magnetic properties of the Sun.

Sensitive global telescope networks that enable frequent or continuous monitoring at optical and infrared wavelengths are needed to resolve stellar behavior over a wide range of time scales. These include stellar flares over seconds to minutes; stellar rotation, and magnetic spots or cloud structures over hours to days; and binary orbits and supernova progenitor and post-explosion evolution over months to years. These networks are also critical for prompt study of transient events, including mergers, tidal disruptions, SNe/GRBs, and gravitational wave events.

These observational capabilities have to be matched with investment in laboratory capabilities to provide the necessary atomic data to characterize highly ionized metal atoms, molecular opacity data at low temperatures, and pressures spanning the ISM to cool stellar atmospheres. As the solar metallicity problem shows, interpretation of stellar spectra is subject to systematic errors, and a thorough study of atomic transitions under different densities and temperature is needed to resolve these issues. Also needed are studies to model the equations of state of stellar interiors, particularly degenerate stellar interiors.

Investment is also needed to advance high-performance computing for 3D modeling of stellar interiors, atmospheres, and binary-star evolution, and to process and perform real-time analysis of the petabytes-per-day data flows anticipated in future surveys.

Maximizing future science capabilities and outcomes goes beyond investment in facilities; theoretical and numerical studies are also needed to advance our understanding. The next decade of astronomical research will rely on advanced software tools to analyze and interpret massive observational and theoretical data sets. The Image Reduction and Analysis Facility (IRAF) and Astronomical Image Processing System (AIPS) have respectively served the OIR and radio communities well over the past 40 years, but some tools, like IRAF, are not supported any more, and the community-based, mainly unfunded, efforts to make equivalent python packages (PyRAF, Astropy) still lack critical functionality, largely owing to lack of dedicated support from the funding agencies. Similar funding support is needed for development of codes for numerical simulations.

There need to be improvements in open-access computing facilities to enable computationally intensive analysis by the broader research community. Current models of access to national supercomputers are limited to extremely computationally intensive theoretical work. Such facilities are not available for intensive-data analysis.

But perhaps the most important developments needed are long-term archives of astronomical data. The current model for funding archives is haphazard, and very often the future of data of discontinued missions is unknown. There needs to be a system in place that can archive the data to expand their utility over time and scale, including maintaining the archive even after a program discontinues. Also important are open data policies. Astrophysics missions such as Kepler and TESS, and ground-based observatories such as Global Oscillations Network Group (GONG), have shown how open access policies significantly increase the scientific output of research investment. For maximum scientific return, it is necessary for other publicly supported facilities, whether ground- or space-based, to make their data public in a reasonable amount of time.

BOX G.1
Science Questions and Discovery Area

Priority questions:

G-Q1: What are the most extreme stars and stellar populations?

G-Q2: How does multiplicity affect the way a star lives and dies?

G-Q3: What would stars look like if we could view them like we do the Sun?

G-Q4: How do the Sun and other stars create space weather?

Discovery area:

G-DA: “Industrial-scale” spectroscopy

`

TABLE G.1 Summary

Capability	**Science Enabled**	**Future Needs**
Time-domain optical observations of the Sun	(G-Q3, G-Q4) Evolution of solar magnetic structures and related internal changes	(1) High-cadence, high-resolution, continuous monitoring of the Sun with: (a) magnetograms, preferably vector magnetograms, cadence better than 1 per hour; (b) full-disk Doppler, multiple wavelengths, cadence at least 1 per minute; (c) full-disk intensity maps, cadence at least 2 per minute.
		(2) Coronagraph to measure coronal magnetic (not just intensity) fields continuously with at least a daily, if not better cadence.
Time-domain radio observations of the Sun	(G-Q3, G-Q4) Evolution of solar magnetic structures and related internal changes	Broadband (< 1 to > 20 GHz) spatially resolved measurements, frequency resolution better than 5 percent, time resolution of ~10 s.
OIR spectroscopy	(G-Q1, G-Q2,G-DA) Precise temperatures, velocities and abundances, and binary orbits	(1) Highly multiplexed (> 1000 fibers) panchromatic spectroscopy on 4–10 m telescopes for sample sizes > 109 stars through the Milky Way in wavelength range from UV cutoff to M band. Spectral resolution $\gtrsim$ 20,000 for detailed abundance work and rotational velocities, R ~ 2000 for bulk abundance and radial velocities.
	(G-Q1,G-Q2,G-DA) Precise temperatures, velocities and abundances	(2) Reasonable resolution (R > 20,000, R ~ 45,000 ideal) spectrograph on ELT class telescope to study stars in Milky Way satellites and Local Group galaxies, as well as the faintest stars in the Milky Way.
	(G-Q1, G-Q2, G-Q3) Brown dwarf characteristics and cloud structure	(3) M band capability on large telescopes.
	(G-Q2, G-Q3,G-Q4) Mass-loss diagnostics for massive stars	(4) Capability for routine Doppler Imaging and magnetic field measurements at very high resolution (R ~ 100,000).
Multi-epoch spectroscopy	(G-Q1, G-Q2) Binary orbits (G-Q3) Brown dwarf cloud structure	(1) Ability to monitor brown dwarfs and low-mass stars over hours to days.
		(2) Survey over months to years with a weekly cadence to detect brown-dwarf binaries. R ~ 10,000–30,000.
		(3) Spectroscopic survey to detect stellar binaries with periods both longer and shorter than Gaia capabilities and beyond the solar neighborhood with multiple observations on 4+ m class telescopes.

Capability	Science Enabled	Future Needs
		(4) Observation of molecular lines (mid-IR to near-IR) for cool massive stars, cadence of months to years.
		(5) Capability of time-resolved observations (cadence days) of transients.
OIR monitoring	(G-Q1) Asteroseismic characterization	(1) A Kepler-like instrument that can monitor one part of the sky for a long time with pixels small enough to avoid confusion; neither TESS not ESA's planned PLATO mission fulfill both conditions.
	(G-Q2, G-Q3, G-Q4) Mass-loss from stars	(2) Targeted optical and NIR spectral observations of narrow emission line evolution in follow-up of events. Time scales of weeks to years.
	(G-Q1–G-Q4) Characterizing hot stellar sources	(3) Near-to-far IR monitoring on years to decades time scales for continuum excess measures on hot stellar sources and episodic changes thereof.
	(G-Q1, G-Q2) Brown dwarf eclipsing binary characterization	(4) Source-by-source continuous monitoring for several hours per source over multiple nights on dedicated 1–2 m+ class ground-based facilities.
UV spectroscopy and spectropolarimetry	(G-Q1, G-Q4) Abundance of the most metal-poor stars (G-Q3) Stellar surface feature mapping (G-Q3) Magnetic field mapping	High-resolution spectrometer and spectropolarimeter covering wide wavelength domains to cover lines formed from 104 to 107 K. High resolution (like those of COS and STIS) for observing abundances of heavy elements in the most metal-poor stars.
X rays	(G-Q3) CME detection, coronal heating, star-exoplanet interaction, mass-loss from evolved stars	Both gratings and microcalorimeters to ensure high resolution (aim for R ~ 5000 to 10,000) and spatial resolution <~1 arcsec over the full range from soft to hard X rays to measure broadest array of charge states.
Radio	(G-Q1) Direct measurement of stellar radii (G-Q3, G-Q4) Brown dwarfs, CMEs, mass loss, interaction of ejecta with ISM	(1) Investment in low-frequency (MHz) facilities.
		(2) Development of a more sensitive VLBI array at frequencies of ~10–20 GHz.
		(3) Radio interferometers with continuous frequency coverage from ~10–400 GHz on sufficiently long baselines (~30–300 km).
		(4) Ability to make repeated observation over weeks to years.
High angular-resolution UV/O/IR/radio imaging and interferometry	(G-Q1) Direct measurements of stellar radii, resolved population studies across the Local Volume (G-Q1, G-Q2) Resolved astrometric and	(1) Maintaining and expanding the leading U.S. capabilities in optical long-baseline interferometry, such as the CHARA Array, NPOI and MROI. Ensuring that sub-mas resolutions are possible. Need baselines of order

Capability	Science Enabled	Future Needs
	binary orbits (G-Q3) Spot modeling	1 km (3× CHARA) with 2–3 m aperture telescopes with 0.9–1 μm instrumentation to provide a full mapping of stellar and brown dwarf radii across the minimum in the hydrogen degenerate mass-radius relation.
		(2) Radio interferometers with continuous frequency coverage from 10–400 GHz on sufficiently long baselines (~30–300 km) to resolve radio photospheres of nearby evolved stars and enable tomography of AGB stars.
Supporting capabilities	(G-Q1–G-Q4)	(1) Improved atomic data for stellar spectral lines and opacities.
		(2) Improved atomic data for highly ionized species.
		(3) Improved molecular opacities and Lande g factors.
		(4) Improvement and standardization of spectroscopic reduction and analysis techniques.
		(5) Standardization in archiving, and broad community access to data products.
		(6) Development of novel approaches to explore the highly multidimensional data sets.
		(7) Improvement in computational power in facilities such as the Long Wavelength Array to enable imaging of the full sky at wide bandwidths continuously.
		(8) Mechanism to enable simultaneous measurements in radio, OIR, and UV facilities.
Theory and modeling	(G-Q2)	(1) Models of binary-star evolution with rotation.
	(G-Q3, G-Q4)	(2) Multidimensional stellar interior and atmosphere modeling.
	(G-Q3, G-Q4)	(3) Multidimensional stellar dynamo models.
	(G-Q2, G-Q3, G-Q4)	(4) Models of/with mass loss.

H
Report of the Panel on an Enabling Foundation for Research

H.1 INTRODUCTION

The Astro2020 steering committee charged this panel with the following tasks: *(1) summarize the current state of resources and support, (2) identify major challenges, and (3) make suggestions to the Astro2020 committee on the topics of computation, simulation, data collection, and data handling; funding models and programs; laboratory astrophysics; and general technology development programs.*[1]

To address its charge, the panel relied on the many valuable white papers submitted by the scientific community, presentations at its three panel meetings, previous National Academies of Sciences, Engineering, and Medicine studies, interactions with other science and program panels, and member expertise. The sections on Explorers and mid-scale projects were based on work of interpanel task forces that drew from the other relevant prioritization panels.

H.2 INCREASING THE INVESTMENT IN THE ENABLING FOUNDATION

People are the enabling foundation of scientific advancement. The key outcome from many of the programs that are the subject of this report, ranging from the theory programs to the suborbital program, is the development, training, and support of scientists. By investing in programs that enable people with the broadest possible range of backgrounds to contribute to scientific advancements, reducing barriers to entry, and providing access to state-of-the-art tools, training, and facilities, the profession's scientific productivity will be maximized. It is people who are the source of U.S. scientific and technical prowess.

In the long history of the decadal surveys, this is the first panel to be explicitly charged with focusing on the "enabling foundation." However, previous decadal surveys have discussed many of the issues raised in this report. Astro2010 and the subsequent midterm assessment emphasized the important opportunities enabled by the Explorer program and mid-scale programs at the National Science Foundation (NSF). Astro2010 stressed the importance of investments in theory, technology development, and laboratory astrophysics. However, many of the augmentations in funding for these programs recommended in Astro2010 were not realized. This lack of investment in the enabling foundation limits the profession's ability to reap the benefits of its investments in telescopes and delays the development of key technologies.

Over the past decade, there has been significant growth in investment in instrumentation and mission capabilities. While the NASA astrophysics budget has grown by 40 percent from fiscal year (FY) 2010 to FY 2020, the investment in people through the NASA research portfolio comprising the Astrophysics Research and Analysis Program (APRA), Astrophysics Theory Program (ATP), Astrophysics Data Analysis Program (ADAP), and Exoplanet Research Program (XRP) has grown only 17 percent, slower than the consumer price index and much slower than the growth of costs at universities and research institutes. With the number of proposals nearly doubling over the decade, the success rate and the inflation-adjusted funding levels of grants (i.e., the investment and support of people) have

[1] See Appendix A for the overall Astro2020 statement of task, for the set of panel descriptions that define the panels' tasks, and for additional instructions given to the panels by the steering committee.

declined. The success rate for NASA proposals during the period FY 2003 to FY 2010 averaged 28 percent. This rate declined to 22 percent from FY 2011 to FY 2019. Meanwhile, the absolute funding per grant grew modestly from $496,000 to $570,000 over this period, somewhat slower than the consumer price index and significantly slower than the cost of graduate school tuitions. The combination of current underfunding and significant opportunities in the coming decade motivates the panel's suggestions to refocus investments on enabling foundations. The panel suggests that a 20 percent increase in the funding (above inflation) for these programs would restore the success rate to historical levels in support of researchers, students, and postdoctorates and match the overall growth in the astrophysics program.

At NSF, there has also been a significant investment in instrumentation through both the Division of Astronomical Sciences and the Major Research Equipment and Facilities Construction (MREFC) program. The Vera C. Rubin Observatory, Atacama Large Millimeter/Submillimeter Array (ALMA), and Daniel K. Inouye Solar Telescope (DKIST) will produce a flood of data that requires increased investment in the enabling foundation. Between 2010 and 2017, the average Astronomy and Astrophysics Research Grants (AAG) success rate was 18.3 percent, significantly lower than the physics program and lower than AAG success rates in previous decades. NSF did not provide information on the success rates for the past several years for the field as a whole nor for different demographic groups or subfields. More recent information on success rates and funding levels would have better informed this report.

The low success rate for proposals has resulted in members of the community operating under extreme stress. This has resulted in significant long-term consequences that inhibit the community's ability to accomplish its goals and to retain talent. The obvious fact is that with proposal success rates so low, outstanding people and teams proposing to do outstanding science are not funded. The National Science Foundation's Astronomy and Astrophysics Advisory Committee (AAAC) identified the threat of the falling success rates in its 2015 report and warned of the risk of a runaway effect with researchers constantly resubmitting as the success rates for proposals dropped.[2] Another AAAC study[3] suggests an optimal success rate of 30–35 percent, far above the current rate for the AAG program.

Researchers at all stages of their careers are unable to take full advantage of the massive investments in large infrastructure projects, which puts U.S. researchers at a disadvantage compared with their peers in other countries. The barrier to entry is especially damaging to young researchers whose highly ranked proposals might get turned down multiple times during their critical first years. People from underrepresented groups are also disproportionately affected by these low success rates.[4]

Improvements can be made in the way grants are awarded. For example, in the NSF AAG program, a typical 3-year grant will fund a graduate student and, perhaps, 1 month of summer salary. With a 3-year cadence and an 18 percent success rate, it is essentially impossible to support a student through the completion of a Ph.D. The grants do not support any costs that are critical to lowering the barrier to entry for women and people from underrepresented groups, such as child care and moving expenses. *The panel suggests that the funding agencies could encourage parental leave by requiring that fringe benefits have some minimum benefits, including leave. These stresses and exclusionary practices are limiting the current and future diversity, vibrancy, and productivity of the field.* These new costs to institutions could be absorbed by adjustments in the overhead rate agreements.

Other fields are, of course, concerned about their grants programs as well. A recent *Report of the 2019 Committee of Visitors—Division of Physics*[5] for NSF detailed the state of the grants programs in a number of areas. It paints a picture of a healthy and vibrant community. Focusing on the grants program

---

[2] AAAC, 2016, *Report of the Astronomy and Astrophysics Advisory Committee*, NSF, Arlington, VA, https://www.nsf.gov/mps/ast/aaac/reports/annual/AAAC_2015-16_Report.pdf.

[3] P. Cushman, J.T. Hoeksema, C. Kouveliotou, J. Lowenthal, B. Peterson, K.G. Stassun, and T. von Hippel, 2015, "Impact of declining proposal success rates on scientific productivity," https://arxiv.org/abs/1510.01647.

[4] See Appendix N, "Report of the Panel on State of the Profession and Societal Impacts," for detailed discussion of barriers to entry for underrepresented groups.

[5] NSF, 2019, *Report of the 2019 Committee of Visitors*, NSF Committee of Visitors, Arlington, VA, https://www.nsf.gov/mps/advisory/covdocs/PHY_2019_COV_final_report.pdf.

in the Division of Physics (PHY) offers an unsettlingly stark contrast to astronomy. A few examples include

- Over the course of 4 years, 24 out of 38 proposals by new young investigators were funded in the Gravitational Physics program.
- In Quantum Information Science (QIS) the Atomic, Molecular, and Optical (AMO)-Experimental/AMO-Theoretical/QIS programs have success rates of 47 percent, 35 percent, and 30 percent, respectively. The report went on to explain that proposal pressure was to blame for the QIS program being "lower than the others."
- Elementary Particle Physics "has success rates sometimes as low as 25 percent … and turns away excellent, fund-if-possible proposals."
- In Nuclear Physics, "The funding success rates are approximately 45 percent for Theory and 38 percent for Experiment." While the report concedes that these numbers "may sound relatively high," it goes on to argue the researchers are underfunded and strong proposals were declined.

As astronomy moves to an era of large instruments, the comparison to physics is appropriate. Both have major investments in common infrastructure and both have large and active theoretical and experimental communities. While the average and median grant sizes in the Division of Astronomical Sciences (AST) and PHY are almost identical (~$600,000 and $350,000), there are almost twice as many awarded in PHY. Given the disparity in the success rates, it is relatively easy to argue that the astronomy grants program is underfunded in an absolute sense. NSF physics invests 68 percent of its budget in research and 35 percent of its budget in facilities. NSF astronomy invests 21 percent of its budget in research and 78 percent of its budget in facilities. These numbers do not include the large MREFC investments in the Rubin Observatory and DKIST. In summary, the astronomy grants program is underfunded both in an absolute sense, and relative to physics.

While the NSF budgets for AST and PHY are almost identical, PHY partners with the Department of Energy (DOE), which provides significant support for some major facilities, allowing for more support of the research program. NSF AST supports most to all of the major facilities used by its researchers, resulting in significantly less resources available for the research program. Recognizing the role that AST plays in supporting facilities, the panel suggests that NSF consider a significant augmentation of the AST program in support of research.

AST's large and growing investment in facilities and declining funding for research grants makes it an outlier among NSF directorates. A recent NSB study[6] showed that the fraction of funding for facilities remained flat over the past 20 years for the Division of Materials Research at 10 percent, Division of Physics at 20–30 percent, and the Directorate of Geosciences at 30 percent. AST devoted 60 percent of its budget to facilities in 2002–2016, and this fraction is projected to grow to 80 percent by 2023.

During the past decade, the international community has constructed (or begun construction on) a series of major observatories that are poised to deliver potentially transformative observations: the JWST, ALMA, the Rubin Observatory, Euclid Telescope, DKIST, and the Nancy Grace Roman Space Telescope. These telescopes will begin producing exabytes of data. However, the data alone are not enough. A significant investment is needed in the people who will develop the theoretical framework, build the archives, develop the software, train the community on the new products and tools, create and implement the computational methods, make the vital laboratory measurements, and analyze the data to produce these transformative results. While support for operations and science analysis are part of the budget for large projects at NASA and DOE, these costs are not part of project budgets at NSF.

---

[6] NSF, 2018, *Study of Operations and Maintenance Costs for NSF Facilities*, NSB-2018-17, NSF National Science Board, Arlington, VA, https://www.nsf.gov/pubs/2018/nsb201817/nsb201817.pdf.

The panel suggests that any large investment in new facilities include resources for operations, a data pipeline, supporting theoretical work, and analysis from the inception of the project. This support will grow the research community using these facilities and enhance their scientific return.

DKIST, ALMA, and the Rubin Observatory are very large investments. As noted in the National Academies mid-decadal review:[7]

> NSF and the National Science Board should consider actions that would preserve the ability of the astronomical community to fully exploit the Foundation's capital investments in ALMA, DKIST, LSST, and other facilities. Without such action, the community will be unable to do so because at current budget levels the anticipated facilities operations costs are not consistent with the program balance that ensures scientific productivity.

This passage was quoted in a recent National Science Board (NSB) study.[8] *The panel suggests as guidance that the annual budget of the grants programs be augmented by at least 1 percent of the cost of the construction cost of a space-based project, roughly what is currently done for flagships such as JWST and the Roman Telescope, and 2 percent of the construction costs of a ground-based project. The larger percentage reflects the lower costs of construction on the ground. The panel also suggests that operations (including software support) be part of the MREFC budget for these projects.*

H.2.1 Investment in Archives and Joint Analyses

Archiving centers will continue to provide a critical enabling infrastructure for collecting, curating, documenting, providing community training, and making the data sets accessible. These tasks will be even more essential in the coming decade.

Archives that reliably catalogue events and objects in the sky have always formed a foundation for astronomical research. In the modern era, these archives are digitally curated in databases, and the past, currently ongoing, and future missions from space and the ground across all wavelengths will generate on the order of 500 petabytes by the end of 2030, several orders of magnitude more astronomical data than has been collected in human history. The simulations needed to interpret these data will generate comparable sized data sets. Although the data volume is dominated by a small number of major missions and surveys, the small- and mid-scale data sets are challenging enough in terms of scale and complexity to also require special attention to their archiving needs. In the past decade, the majority of scientific papers based on data from large missions and surveys are archival analyses. In the coming decade, these archival analyses will become even more important and their technical implementation more challenging.

With the missions and surveys planned for the 2020s and beyond, enormous opportunity exists to greatly multiply the scientific return beyond their core goals, using them to address a broader set of both foreseen and unforeseen questions. Illustrative examples of these opportunities include the following:

- Combining images of the sky at the level of the pixels of the maps produced for the major ground and space imaging programs, including but not limited to the Rubin Observatory, the Euclid satellite, the Roman Telescope, the eROSITA mission, and future cosmic microwave background (CMB) experiments, which would allow powerful new constraints based on gravitational lensing at both cosmological and galactic scales, photometric redshifts, galaxy evolution, and motions in the solar system and the Milky Way. This analysis will require

[7] National Academies of Sciences, Engineering, and Medicine, 2016, *New Worlds, New Horizons: A Midterm Assessment*, The National Academies Press, Washington, DC, p. 8.

[8] NSF, 2018, *Study of Operations and Maintenance Costs for NSF Facilities*, NSB-2018-17, NSF NSB, Arlington, VA, https://www.nsf.gov/pubs/2018/nsb201817/nsb201817.pdf, p. 22.

investments beyond those of individual projects to enable this enhanced science return. High-quality cosmological simulations will also be needed to enable these joint analyses.

- Combination of time-resolved and target-of-opportunity observations across facilities, large and small, and wavelengths. If there exists a close enough integration of access methods, formats, and meta-data to efficiently and properly analyze the joint data set, this would allow new windows on transient and variable phenomena.
- Timely analyses of multi-messenger phenomena (across electromagnetic, gravitational wave, and particle detectors), which would allow tests of fundamental physics, the nature of black holes, and cosmology, and triggering of follow-up observations, if they can be performed rapidly and robustly.

The panel outlines a framework for structuring the U.S. astronomical archiving system so that it addresses astronomy's most compelling science goals.

The existing federally funded archiving centers are essential. Operating from the Space Telescope Science Institute (STScI), NOIRLab, the Infrared Processing and Analysis Center (IPAC), the National Radio Astronomy Observatory (NRAO), the Smithsonian Astrophysical Observatory (SAO), the Gravitational Wave Open Science Center (GWOSC), and other locations, they focus on different types of data and have developed unique expertise. The combination of their data, meta-data, documentation holdings, the access tools they have developed, and most importantly, their scientific and technical staff is critical to maximize the scientific return from the profession's investments in new telescopes. Nevertheless, the panel's use cases illustrate that separating archives by wavelength, mission, and funding agency limits the science. Because they are funded by different agencies, including NASA, NSF, and DOE, with different policies and science goals, the coordination necessary to enable new joint capabilities is hampered.

Technological developments deriving from the commercial hardware and software industry are also emerging that can enhance the profession's approach to archives. Today, most astronomical data centers operate their own hardware; however, cloud services (both computing and storage) are an increasingly affordable and flexible way of handling astronomical data. A decade ago, the cutting-edge archives allowed complex server-side queries (e.g., in Structured Query Language [SQL]), which typically would be followed by a data download and local analysis by the end user. Today, science platforms exist or are being developed at most centers (most often relying on Jupyter notebook deployment) to provide more complete server-side analysis tools so that almost all analyses can be performed on the server, with limited if any data download to the end user. A key technological tool behind these science platforms is the "container," an encapsulation of the full operating system of a system that can be duplicated and operated on any hardware.

The proliferation of open source software and software development resources has expanded the suite of tools available to astronomers for building, accessing, and analyzing data from archives. These developments over the past decade have swept into astronomy from the wider world of software and data, and archiving centers are individually preparing to take advantage of these and future, unforeseen developments as they arise in the next decade. Nevertheless, the archive centers have limited ability to adopt these new technological developments in a coordinated and synergistic manner. Insufficient resources are focused toward very few funded efforts in building shared infrastructure among them, again partly owing to their separation across NASA, NSF, and DOE.

A major challenge facing the U.S. astronomical archive system is the need to coordinate its federated archives to address the science needs of the 2020s. The development of standard protocols and tools to implement them over the past 15 years has been a necessary but not sufficient effort to enable this coordination. The panel suggests a more proactive and robust approach is needed to maximize the scientific return from the profession's investments in instruments, telescopes, and satellites.

To face this challenge, the panel envisions an Astronomical Data Archiving System (ADAS) to serve as an umbrella organization to coordinate the activities of the existing archive centers, with the goals of increasing their effectiveness in achieving their missions and opening up new opportunities that

would otherwise be impossible. Models in other fields such as earth science, in the international astronomical community, and in previous and ongoing U.S. virtual observatory efforts can provide lessons in addressing the growing needs and remaining challenges in astronomical archiving. Critical considerations in designing this new coordinated system would include the following:

- Preserving the roles, responsibilities, expertise, and funding streams that define current centers and missions.
- Enabling career paths for archive scientists that allow them to move within the ADAS "family."
- Providing the resources and mission to initiate and lead community-wide efforts focused on broadening participation through education, training, citizen science, and curriculum development in computational skills, software development, and data science.
- Providing the resources and mission to pursue opportunities to develop common resources and shared expertise across centers.
- Enabling science-driven efforts to perform analyses across boundaries of missions and centers.
- Providing support for users to access the simulations needed to interpret the data;
- Having the capability to bring new centers under the ADAS's umbrella and to coordinate/collaborate with international partners.
- Incorporating data from smaller projects that lack the support of a major center and provide support for data archiving/preservation for these projects.
- Providing mechanisms for the U.S. astronomical community to contribute input to the planning and prioritization of ADAS activities.
- Continuing to support journals and access to journals. The NASA Astrophysics Data System, which provides free bibliographic access to 13 million publications records, plays an important role in astronomical research, and will continue to be essential in the coming decade.
- Developing common policies for data storage and transfer specifications.

With funding outside that of the existing individual centers, ADAS could provide the resources to fulfill its mission to make all centers more interoperable and to enable cross-center scientific analysis and to coordinate complementary training programs.

To enable discovery in the 2020s, the panel suggests that the ADAS, the archive centers, the funding agencies, the major NASA-funded missions, and NSF-funded and DOE-funded projects address numerous other challenges through funding of a number of data-related efforts:

- Science-ready data products and APIs could be built into the funding of all new missions and projects, utilizing the available common infrastructure and protocols of the ADAS and/or the archive centers.
- Support for software infrastructure efforts specific to astrophysics, both community-led and those led by archive centers.
- Initiation of new programs and/or supporting community-led efforts in training and education, enabling contributions from a broader range of scientists.
- Preservation of data from smaller or nonfederal projects.
- Standard, machine-readable methods for preservation of meta-data and documentation.
- Direct support for ensuring that important analyses preserve their replicability, through providing needed support for the software engineering effort necessary to do so.

The archiving of astrophysical simulations of theoretical predictions deserves special mention here as an important challenge to which the archives can contribute. "Simulations" includes both

instrument simulation given the astronomical input (e.g., the image of the sky) and the astrophysics simulation of physical laws that might produce a specific image of the sky under some theoretical framework. Here, the focus is on simulations that are essential for statistical analyses of the observations. The panel suggests that archive centers continue to develop the software and tools to allow the instrument simulations of instrument behavior. As both simulations and observations grow in scale, the ability to co-locate observational data with astrophysical simulations will prove necessary, and the panel suggests that archives develop partnerships with the high-performance computing centers and simulation groups necessary to provide this service. The panel suggests that archives of simulations ensure that software for all curated simulations is versioned, traceable, and can be used to replicate the simulation if necessary.

The new data sets from astronomical facilities at all scales in the 2020s will lead to numerous new discoveries and breakthroughs in understanding of astrophysics and physics. The breadth and depth of this new science, how broad and inclusive the community is that contributes to it, and the profession's ability to take advantage of new and unexpected opportunities depends on a well-funded and well-coordinated archiving system designed for the coming decade and beyond.

This suggestion is in line with the recommendations made in the NASA *Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science 2019–2024*[9] document. Indeed, the panel envisions that this new system will address the Open Data/Open Software, High-End Computing, and Advanced Capabilities areas in addition to the items included in the Archives Modernization area. NASA's strategy could be enhanced by a stronger emphasis on community education.

A detailed study incorporating input from the astronomy community and from the existing centers will be needed to appropriately scope and define this system. The panel envisions that the system would require very roughly 50 full-time employees (FTEs), which would include astronomers, software engineers, and other staff in proportions that require a focused future study. This system would add ~$10 million a year in operating costs above the costs of the existing archives. The panel suggests that this be a supplement to existing programs.

This preliminary estimate is based on considering the number of FTEs working on data management, processing, and distribution at the major current and planned facilities, which add up to hundreds of FTEs. To coordinate these systems in an effective manner requires a sufficiently large investment of central effort, which motivates the considerable investment envisioned here (although fractionally this investment is smaller than Earth Science Data and Information System (ESDIS), a similar system in Earth Sciences). It is essential that, irrespective of the goals laid out above, the ultimate scope of activity and the expectations of any ADAS-like system be tuned appropriately to its available resources. While some of the ADAS FTEs would be located at the existing larger centers, some may be best co-located with smaller projects. *As discussed below, the panel suggests considering an approach where NASA serves as the lead agency for an interagency supported archiving program, and NSF and DOE are the lead agencies for providing access to high-performance computing resources to the broader national community.*

H.2.2 Software

Software development is an essential part of almost all aspects of astronomy, and software developers, perhaps better called "software instrument builders," are an essential part of the astronomy community. However, neither are sufficiently funded or supported by existing structures.

The profession has entered an era in which the ultimate success and impact of major programs will be equally dependent on software and hardware development. As such, software development needs

[9] NASA, 2019, *Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science 2019–2024*, Washington, DC, https://science.nasa.gov/science-red/s3fs-public/atoms/files/SDMWG_Full%20Document_v3.pdf.

to be included in budgeting, planning, and career development. In this panel report, software includes data reduction pipelines such as Astropy and analysis packages developed by large projects, large software projects such as Modules for Experiments in Stellar Astrophysics (MESA)[10] or Enzo,[11] and codes used by individuals or small groups to produce results in published papers.

Software development has evolved significantly over the past several decades. Large teams of people often work together to write and develop software. For high-performance computing, these large teams are essential in part because of the increasing complexity of the problems being solved, and in part because of the increasing complexity of computer hardware. Today's astronomical software must take advantage of heterogeneous computing systems currently available, from graphical processing units (GPUs), to multi-core processors running in collections of thousands of nodes, to standard general-purpose CPUs running on laptops. For both theoretical models and data analyses, the complexities of the systems modeled demands large codes developed by teams with a wide range of expertise. The practices for developing software as a team are quite different from developing software as an individual. The panel envisions that future astronomical training will include best practices for developing code as part of a larger team and for a diverse range of computing hardware.

The Image Reduction and Analysis Facility (IRAF),[12] developed and supported through federal funding for three decades through multiple avenues, demonstrated how a workhorse software system, freely available, could be immensely empowering to the community to make use of all types of data from many types of instruments and telescopes. However, the lack of federal support for the modernization of this stalwart resulted in a gaping hole in the software infrastructure available for data reduction and analysis. This hole has been partially filled by the mostly volunteer-run Astropy Project and the Astropy package. However, the Astropy Project may need to abandon its support and development of its community software, in the same way the IRAF project was eventually forced to do, because of the lack of sufficient and reliable funding. These examples are illustrative of how the normal grant funding structure works poorly in the context of building software infrastructure intended to undergird most modern astronomical software. Any effort to create infrastructure like that depends on continuous fundraising efforts with short time horizons, without any path to earning the longer term commitments that are necessary for longer term planning and stability. These infrastructure development efforts require reliable federal funding along with the software and the people maintaining it.

By investing in software training for the entire astronomy community and in astronomical software developers, the profession can build the tools for transformational astronomy in the coming decade. This will require support for making code open source and the maintenance of large codes being developed by both individuals and by broad community efforts. These investments will improve reproducibility and reduce unnecessary duplication of codes.

Astronomical software development is training a generation of people who are finding exciting opportunities outside astronomy. Without more funding opportunities and career tracks within astronomy, it is challenging to retain these software builders.

The 2018 National Academies report *Open Source Software Policy Options for NASA Earth and Space Sciences*[13] makes important observations:

> SMD [Science Mission Directorate] needs to foster a new culture of openness and encourage a social norm of sharing and collaboration, in part by incentivizing the development of OSS [Open Source Software] in the academic community through the use of targeted grants, fellowships, and prizes. The move toward openness is also facilitated by the establishment and use of open source

[10] MESA, "MESA home," http://mesa.sourceforge.net, accessed July 26, 2021.
[11] The Enzo Project, "Home," https://enzo-project.org, accessed July 26, 2021.
[12] IRAF Community Distribution, "Home," https://iraf-community.github.io/, accessed July 26, 2021.
[13] National Academies of Sciences, Engineering, and Medicine, 2018, *Open Source Software Policy Options for NASA Earth and Space Sciences*, Washington, DC, The National Academies Press, p. 3.

libraries (code and tools used by programmers when writing software) to collect and disseminate community software.

The 2020 NASA Research Opportunities in Space and Earth Science (ROSES) call[14] is responsive to this part of the 2018 National Academies report through its funding of open source software and open source tools, frameworks, and libraries. However, with funding at a modest level, this is only a small step toward supporting these programs.

The National Academies report suggests the possibility of requiring software management plans for all proposals. The report further suggests that proposals that involve releasing software include a budget description for software development, documentation, distribution, support, publications, and maintenance. The addition of software maintenance plans in proposals encourages open code and longer term code maintenance by the code authors themselves and allows the peer review process to set the pace of culture change. While there are costs associated with maintenance, there are also significant savings in having the continued usability of code developed at high cost.

Hardware and software environments are constantly evolving, and in the face of this changing environment the astronomical software community needs to embrace sustainable software solutions. "Containerization," for example, is currently an effective solution for codes to remain operable. Containerization is an encapsulation of the full operating system so that the code can be operated on any hardware. The panel has identified the implementation of containerization as one of the goals of the ADAS described above.

Replicability and reproducibility is an essential part of the scientific process. For software, this requires supporting, incentivizing, and educating the community about best practices that preserve the software and input files needed to reproduce and replicate analyses and results. This will likely need to involve coordinated efforts among investigators, collaborations, science libraries, publishers, archives, repositories, and the federal agencies. The American Astronomical Society (AAS) journals publication of software papers and partnership with the Journal of Open Source Software, the availability of Github repositories, and the European Zenodo repository are all encouraging developments. Improving standards for citing software[15] encourage proper crediting of work, document the ingredients used in an analysis, and are essential to enabling replicability and reproducibility of results.

NASA's Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science 2019–2024 presents a forward-looking vision in this area. The NSF Division of Astronomical Sciences (AST) could collaborate with the Computer and Information Science and Engineering division (CISE) to develop its strategic plans for supporting astrophysics for software development and data management. Similarly, DOE Cosmic Frontiers could work with the Advanced Scientific Computing Research (ASCR) program to develop a strategy for its astrophysics projects. These strategies could include training in software engineering, computer science, and programming practices for the astronomical software development community.

H.2.3 Theoretical Astrophysics

Theory often drives fundamental new discoveries, as well as informing the design and operation of new observations. Support for both individual investigators and theory networks through the agency

---

[14] NASA, 2020, "Research Opportunities in Space and Earth Sciences—2020 (ROSES-2020)," NASA HQ, Washington, DC, https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=735966/solicitationId=%7BBCEE336B-D550-CCBA-1C8C-7A866DB06F45%7D/viewSolicitationDocument=1/FULL%20ROSES-2020_Amend45.pdf.

[15] D. Bouquin, G. Muench, K. Cruz, D. Chivvis, and E. Henneken, 2019, "Citing Astronomy Software: Inline Text Examples," AstroBetter, https://www.astrobetter.com/blog/2019/07/01/citing-astronomy-software-inline-text-examples/.

grant programs is essential for the health of theory programs; however, funding levels have remained flat and proposal success rates have been dropping throughout the past decade.

Theoretical astrophysics has grown to encompass analysis and interpretation of data, computational methods that enable investigation of complex physical systems, as well as the more traditional pencil-and-paper calculations, a growth that at times blurs the boundary between theory and observation. Theory has long played an outsized role in discovery in astrophysics. In some cases, entirely new observational programs have been developed to test theoretical predictions, often with spectacular success. For example, measurements of fluctuations in the cosmic microwave background radiation have provided remarkable new insights into fundamental properties of the universe. Theory is crucial for the interpretation of most observations—for example, the theory of stellar structure for asteroseismology data, or orbital mechanics for exoplanetary systems. Last, theory is used to develop essential new tools and frameworks for analysis of large and complex data sets, especially those resulting from surveys.

In recognition of the important role of theory, the Astro2010 decadal survey recommended a new approach for support involving augmentation of existing grants programs and the creation of new ones. A modest ($8 million per year) augmentation was recommended for the NSF Astronomy and Astrophysics Research Grants (AAG) program, which supports investigators in all areas of astronomy including theory.

An additional important recommendation for theory in the Astro2010 survey was the creation of a new interagency funding opportunity for Theory and Computation Networks (TCAN). This program was in recognition of the increasingly complex nature of modern research problems, and the increasing reliance on ever more complex software and numerical methods for their solution. The program was initially proposed to support 5-year grants jointly funded by all three agencies, but the DOE declined to participate, owing to their already strong support of computing through, for example, the INCITE program. Both NSF and NASA did initiate a TCAN program,[16] but only for 3-year projects, and NSF discontinued the program after only one solicitation.

For the future, several issues remain important. The NSF AAG program is a crucial vehicle for funding new independent and novel investigations in astronomy and astrophysics, but especially in theory. The panel suggests that this program continue to be supported in the face of continuing budget pressures.

Second, the original intent of the TCAN program was to promote cross-agency and cross-disciplinary teams to tackle challenging problems. The implementation of the program, while providing welcome support for theory, did not achieve this vision. Some of the factors that contributed to this lack of success included cross-agency support for the program lasting only 1 year, with NASA left to solely support the program in all subsequent years, and funding for the program being shifted from other programs, rather than being allocated as new funding.

If there is an opportunity for new initiatives in the future, a revived TCAN program that truly promotes interaction across agencies (involving, e.g., both ground- and space-based facilities), and across disciplines (including astrophysics, applied mathematics, computer and data science, physics, etc.), could have significant impact.

Last, there remain concerns regarding barriers to funding cross-disciplinary projects such as large code development projects (involving, for example, computer science and astronomy), and multi-messenger astronomy (involving physics and astronomy).

A significant (25 percent) increase in the funding for the NASA Astrophysics Theory Program (ATP) was recommended by the Astro2010 survey, but unfortunately this also was not realized. Instead, not only did funding remain flat, but also the program moved to a 2-year cycle of proposal. This cadence negatively impacts career development across the community. The panel suggests that the agency implement at least this augmentation of 25 percent and that the program resume its annual cadence.

[16] NSF and NASA, 2013, "Theoretical and Computation Astrophysics Networks (TCAN)," NSF 13-512, https://www.nsf.gov/pubs/2013/nsf13512/nsf13512.htm.

H.2.4 High-Performance and High-Throughput Computing

Computation, both in theory and in data science, has emerged as foundational for essentially every topic in astronomy and astrophysics. Increasing capability in computing infrastructure is crucial for scientific progress. DOE and NSF are significantly increasing their capacity in high-performance computing over the coming decade, while NASA is expanding at a slower rate.

High-performance computing (HPC) involves undertaking detailed calculations at high speeds using large supercomputers. HPC tasks require a large amount of compute power for solving complex problems (e.g., simulations of physical processes).

High-throughput computing (HTC) enables relatively simple computational tasks to be undertaken in a highly efficient way (e.g., processing and analysis of very large data sets).

In the past decade, computation, both in theory and data science, has unarguably emerged as foundational for essentially every topic in astronomy and astrophysics. Numerical simulations and big data analysis have become increasingly sophisticated, and their role in astrophysics has correspondingly experienced enormous growth. Despite approaching the limit of Moore's law, computational power has also been growing steadily, with exascale supercomputers expected to become publicly available as early as 2022, and the potential for significant expansion in these capabilities by 2030. With the increasing sophistication in software, analysis, and computational capability, there is enormous potential and opportunity for scientific discovery in the coming decade.

HPC enables discovery through simulations of processes such as the formation and evolution of stars, planets, galaxies, the universe, and gravitational wave events. HTC enables discovery using large data sets including investigations using archived observational data, joint pixel processing of complementary observations, and analysis of large, simulated data sets. Continued and expanded support for increasing capability in computing infrastructure and in people with sufficient expertise in HPC and HTC is crucial for scientific progress.

The size of observational and synthetic data sets has consistently increased over time, from terabytes to petabytes, and soon to exabytes. In this era of Big Data, there is an emerging opportunity to use publicly available cloud computing for cost effective solutions, rather than hosting huge hardware resources and numerous proprietary facilities. The utility of cloud computing for HTC is currently clearer than for HPC, although this may be changing. The panel suggests that the funding agencies continue to explore the potential of cloud computing for a range of efforts and to provide support for projects to utilize cloud computing where appropriate.

In the coming decade, DOE and NSF have plans to significantly increase their capacity in HPC, while NASA plans to expand at a slower rate. To ensure sufficient HPC resources for missions and to ensure that the community has sufficient access to HPC facilities over the coming decade will require either coordination among the funding agencies, or an increased expansion of HPC facilities at NASA. However, funding that supports work across agencies is sparse. As noted in Section H.2.3 above, the Astro2010 decadal survey recommended the development of a TCAN program, which was intended to be a collaboration between NASA and DOE for space-based astronomy and NSF and DOE for ground-based astronomy. A TCAN program was initiated between NASA and NSF, but has since become a NASA-only program. The panel suggests that reinvigorating a focused collaboration between all three funding agencies will enable the most efficient use of resources and will facilitate rapid development in key advanced computing areas that are currently experiencing only moderate progress owing to lack of support.

In the past decade, HPC simulations have become integral to theoretical modeling, forecasting and survey formulation, and in the eventual analysis and interpretation of observational data. Developing and exploiting the software to undertake these tasks not only requires specialized facilities with large computing and storage resources, but also people with extensive expertise in both computer science and astrophysics. The potential barriers to access for individual investigators is of great concern to the community, and several white papers were submitted to highlight these concerns, particularly for those at institutes without preexisting relationships with the large HPC and HTC facilities.

The panel suggests that the agencies increase their investment in training in the use of HPC and HTC facilities. In order to ensure appropriate training to undertake HPC and HTC programs and more equitable access to HPC and HTC facilities, it is important that extensive training be available at the undergraduate and graduate level. In addition, the panel suggests that more support be provided for training directly from the facilities—including workshops and internship programs—for all career levels.

Coordination among the funding agencies would benefit the broader national program. The panel suggests an approach that involves NASA taking the lead in supporting archiving, and DOE and NSF taking the lead in providing HPC supercomputing facilities to scientists across astrophysics.

H.2.5 Data Science and Machine Learning

The interaction between astronomy and data science is a fruitful two-way exchange. Data science advances enable new insights in astrophysics. Rich astronomical data sets with underlying physical symmetries can push technology development in data science. Support for ongoing training in new data science techniques will enrich the return from large data sets and advance both data science and astronomy.

Over the past decade, data science has advanced dramatically. Machine learning techniques are playing an increasingly important role in astrophysics, and this trend is likely to continue into the future. Over the past few years, there have been multiple joint data science/astrophysics faculty appointments. Many universities are adding new courses in this field. Both undergraduates and graduate students are pursuing joint degrees in programs that did not exist in 2010.

Astronomical data offers many opportunities for data science research and is already proving to be a valuable data set. For example, Stalzer and Mentzel[17] ranked the Sloan Digital Sky Survey (SDSS) as the sixth most influential paper in Big Data, just behind Shannon's classic information theory. Astronomical data is valuable for data science for many reasons:

- The data is open access, has no commercial value, and is free of the many ethical issues associated with other kinds of image data. In contrast, images of faces scraped from the Internet often do not have permissions, are used for photo surveillance, and are often racially biased samples.
- Astronomical data is rich and ranges from images, tables, graphs, and uneven time series to multi-dimensional grids.
- Data in the physical sciences is structured with individual particles, planets, and stars interacting in particular ways and with well-understood symmetries—data structures that differ from the widely studied images and sequences in other areas. This rich structure has already inspired early work in graph neural networks and geometric deep learning.
- Astrophysicists have high-fidelity simulators that capture mechanistic causal models that describe both the astronomical phenomenon (e.g., the evolution of large-scale structure) and the astronomical processes (e.g., observations of gravitationally lensed galaxies by the Rubin telescope). In recent years, astrophysicists and data scientists have developed numerous new techniques for likelihood-free inference, advances in density estimation, implicit generative models, and probabilistic programming. These techniques are now being used across a wide range of fields (e.g., particle physics, chemistry, and neuroscience) and are part of an emerging new area spanning machine learning and the physical sciences.
- Because it is possible to simulate data, it is possible to query whether a model is overfitting the data. Since the underlying physics is known for many astrophysical data sets, it is possible to learn whether artificial intelligence (AI) is learning the true underlying rules. This is a

[17] M. Stalzer and C. Mentzel, 2016, A preliminary review of influential works in data-driven discovery, *SpringerPlus*, 5:1266, https://doi.org/10.1186/s40064-016-2888-8.

much more significant test of a model than cross-validation and is important for making models safe and actually improving the understanding of science.

- Building theories for the physical world is a potentially less ethically fraught implementation of AI on classifying text or images. While much of the work in data science has been focused on images, because they require no domain knowledge and are useful for online advertising and customer analysis, physical systems may be a better path toward Artificial General Intelligence (AGI).

Data science offers powerful new tools for studying astronomical data and astrophysical systems. Machine learning has already shown significant success as a tool for identifying anomalies in data. These techniques could lead to transformative discoveries from the new data sets expected to become available in the 2020s. Machine learning has the potential of increasing the amount of information obtained from astronomical data sets by enabling modeling of complex nonlinear phenomena and instrumental effects. If machine learning can be successfully used to model multi-scale phenomena, it could open up the ability to more accurately simulate a wide range of astronomical processes from planet formation to galaxy formation.

As this is a rapidly evolving field, the panel suggests that funding agencies use the grant programs and existing data centers to initiate and support ongoing training for astronomers as well as broad opportunities to enable a diverse group of scientists to apply and teach these techniques.

H.2.6 Laboratory Astrophysics

Laboratory astrophysics is essential for enabling science across astrophysics, and new laboratory measurements are essential to realize the full potential of recent and imminent major observatories (ALMA, JWST, GMT/TMT, etc.) targeting stars, planets, star and planet formation, and high-energy phenomena. If the aim is to understand the structure and evolution of stars, galaxies, and the universe as a whole through the observations from future facilities, laboratory astrophysics will be required. Since astronomical systems span an enormous range of densities and temperatures, developments in laboratory astrophysics have the potential to stimulate developments in chemistry and physics.

Currently, there are relatively few groups in the United States that are making the needed laboratory astrophysics measurements, and the prospects for establishing new groups is limited by an overall small funding envelope. This needs to be addressed to maximize the scientific return of the major astronomical investments in the 2020s. To expand the field of laboratory astrophysics and to ensure that existing expertise is transferred to a new generation, the overall funding envelope needs to be increased, and barriers to entry need to be removed.

Despite limited resources, laboratory astrophysics has been instrumental in advancing astrophysical discoveries in the past decade. In the search for our interstellar chemical origins, the 2010s delivered first identifications of aromatic organics, and chiral molecules, and the first inventories of organic molecules at the onset of planet formation. These results were obtained because of new spectroscopic line lists. Complementary laboratory work revealed that many of these organics can form in icy grain mantles at close to 0 K, and that complex, prebiotically interesting organic molecules are thought to be ubiquitous during star and planet formation. New laboratory data has also been key to characterize the atmospheres of exoplanets; experimentally determined molecular line opacities at high temperatures have enabled retrievals of water abundances and constraints on atmospheric carbon/oxygen ratios, while haze formation experiments have been key to elucidate what kind of hazes and clouds may form on different kinds of exoplanets. Laboratory astrophysics has also been instrumental in advancing the fundamental understanding of the underlying physics governing stars. For example, in the early 2010s, a discrepancy between the theoretical convection zone boundary within the Sun and the value implied by asteroseismic data was identified. This became known as the solar convection zone boundary

problem and represents a lack of understanding in the physics that govern the closest star, the Sun. The Sun is the benchmark by which all other stars across the Milky Way and beyond are understood. Recent laboratory measurements of the opacity of iron at the temperature and density of the solar convection zone boundary revealed large discrepancies between the observed and theoretical opacities, implying that theoretical stellar opacity calculations are far from correct. It is because of these laboratory measurements that the solar convection zone boundary problem has been reduced significantly.

Laboratory astrophysics was identified in the Astro2010 decadal report as "vital for optimizing the science return from current and planned facilities,"[18] especially in the ALMA and JWST era. Yet, they found that "support and infrastructure for laboratory astrophysics are eroding both in the National Laboratories and in universities," and they recommended that "the funding through APRA that is aimed at mission-enabling laboratory astrophysics should be augmented at a level recommended by this scientific assessment … a notional budget increment of $20 million over the decade may be required."[19] This augmentation was not implemented. As a result, the community is now in the age of ALMA and (soon) JWST, without many of the required laboratory measurements, which will severely limit the science return from these observatories if not addressed in the 2020s. ALMA and JWST have incredible capabilities to explore the interstellar medium and star and planet formation, and in the case of JWST, to characterize planets. However, interpreting this data requires laboratory and computationally generated databases of dust and molecule opacities, complex-refractive indices of condensed matter particles (aerosol analogs), spectroscopic lines, collisional cross sections, and gas and solid-state reaction rates. At present, these are all woefully incomplete, and there is a small number of active laboratories that contribute to them.

The 2020s will also see a number of large observational surveys focusing on stellar astrophysics, which together address the fundamental astrophysical problem of stellar properties, such as the detailed chemical compositions, masses, and ages. In the era of upcoming photometric (Rubin Observatory, Skymapper, etc.) and large high- to low-resolution spectroscopic surveys (SDSS-IV, SDSS-V, 4MOST, WEAVE, GALAH, Gaia-ESO, etc.), astronomers will not be limited by data in the pursuit of stellar astrophysics, but rather by a lack of laboratory measurements needed to interpret the data. While these fundamental parameters are crucial for stellar astrophysics, they are also important in a wide range of astrophysics ranging from exoplanet science to galaxy formation. The availability of relevant laboratory atomic, molecular, and optical (AMO) data, such as highly accurate wavelengths, transition probabilities, photoionization cross sections, line broadening parameters, and collisional cross sections, will be critical for maximizing the scientific return of these surveys, observatories, and missions, which together represent a significant investment of U.S. astronomy resources.

At higher energies, the scientific return from proposed high-resolution X-ray spectroscopic missions, like Athena, will not be able to capitalize on their high resolution without new atomic data including collisional and photoionization cross sections. Potential diagnostics of density, temperature, ionization, abundances, and so on will not be realized without improved laboratory data on transition energies, electron impact ionization collision strengths, photoexcitation, and ionization. Laboratory astrophysics is also a required foundation to enable science on a range of scales—from as small as dust grain growth to the solar convection boundary problem, to understanding the shock physics of supernovae.

Laboratory astrophysics is mainly funded via grants by NASA Astrophysics Research and Analysis (APRA) and Astrophysics Data Analysis Program (ADAP), and the NSF AST. However, the number of awards approved across all of these programs is small, and they have been declining since the early 2000s. This put the laboratory astrophysics field under severe pressure during a time when there are growing needs from the ALMA and JWST, and stellar and exoplanet astrophysics communities. The

[18] National Research Council, 2010, *New Worlds, New Horizons in Astronomy and Astrophysics*, Washington, DC, The National Academies Press, p. 32.

[19] National Research Council, 2010, *New Worlds, New Horizons in Astronomy and Astrophysics,* Washington, DC, The National Academies Press, pp. 220–221.

panel suggests that going forward it is critical that agencies continue to fund the currently active groups, train the next generation of laboratory astrophysicists, and lower barriers of entry into the field.

The panel suggests increased investment in laboratory astrophysics. This increase could come in many forms, such as (1) increased investment in the laboratory portion of the NASA APRA program; (2) inclusion of a special funding line of laboratory astrophysics in missions and facilities, possibly through NASA Phase E funding and something analogous for NSF-supported facilities; (3) an increase in the funding for the NSF Major Research Instrumentation (MRI) grants and removal of the institutional cost-sharing requirement; and (4) an increase in the funding for interdisciplinary but laboratory astrophysics-centered workshops, internship, and professional development programs at the National Laboratories and other laboratory astrophysics centers around the United States that will interface with universities. While these remedies are not the only possibilities, they represent four clear actions that can be taken to strengthen laboratory astrophysics and enable the United States to make the most of its astronomical investments and achieve the scientific aims of the next decade.

H.2.7 Technology Development

Investment in astrophysical technology development is broadly enabling for astrophysics and enhances U.S. technological competitiveness.

Astrophysical discovery and technological advancement march hand-in-hand, with new technology opening new and unexpected windows on the universe. Consequently, the astrophysical community has a core strategic need to mature new technologies to the point where they can be flown on NASA missions or deployed on ground-based instruments.

Conversely, there is clear synergy between the astrophysical community's needs and expertise, and those of broader society. To seek life on other worlds, astronomers require essentially noiseless, nearly quantum limited detectors in the UV, visible, and IR. Many of these same properties are needed for quantum computing and information science. Robotics, automation, and advanced manufacturing enable building the next generation of telescopes on the ground and in space—and are likewise strategically important to U.S. technological competitiveness. Astronomers routinely contribute to—and draw on—advances in materials science for their electro-optical sensors, optical components, and system engineering to build ultra-precise instruments.

H.2.7.1 NASA

NASA tracks risk using Technology Readiness Levels (TRLs). A TRL is an integer in the interval [1,9], with TRL 1 roughly corresponding to an idea with supporting data up to TRL 9 denoting proven, flight-heritage hardware. All key technologies are generally required to be TRL 6 early in the life of a project. For strategic missions (JWST, Roman, etc.), TRL 6 is required to pass the Preliminary Design Review (PDR). For Explorers, SmallSats, and CubeSats, all technologies are generally required to be TRL 6 or higher upon selection for implementation.

In addition to directed funding, NASA uses two grant programs to support technology development. These are the Astrophysics Research and Analysis (APRA) Program and the Strategic Astrophysics Technology (SAT) Program. APRA is open to all TRLs, whereas SAT focuses on the "mid-TRL" range from 3 to 6.

APRA is broad-based, and not specifically focused on strategic missions. It encompasses Suborbital/Suborbital-Class Investigations, Detector Development, Supporting Technology, and Laboratory Astrophysics. During the period 2014–2018, about 50 percent of new APRA funding went to the Suborbital Program. The remainder was split with about 20 percent each going to Detector Development and Technology Development, and about 10 percent to Laboratory Astrophysics. Although the typical $200,000 to $400,000 per year award amounts for Technology Development are appropriately

sized for individual investigators supporting perhaps a postdoctorate or a few students, in practice APRA awards are often too small to allow investigators to partner with industry. Yet, having industrial partners is essential for developing key enabling technologies including advanced optics and detectors.

The SAT program was created in response to Astro2010 as part of addressing a mid-TRL gap from TRL 3–TRL 6 that was perceived to exist for strategic missions. SAT solicitations are highly responsive to "technology gaps" that strategic missions identify with substantial input from the community in Program Analysis Groups (PAGs). Although the PAGs solicit input from individual investigators, in practice the needs of large, strategic missions hold more sway.

Although the SAT program addresses the needs of large, strategic missions, a notable mid-TRL gap still exists for Explorers, SmallSats, and CubeSats. Typical APRA award amounts are insufficient to support Detector Development or Supporting Technology development by a PI-led team working with the major technology vendors.

The NASA Space Technology Mission Directorate (STMD) has historically provided support for developing long-range technologies relevant to astrophysics. In the past several years, this support has been eliminated. This has led to a decrease in support for challenging long-term technology.

The APRA Program offers the best opportunities for developing highly innovative but risky new technologies. Unfortunately, many excellent proposals cannot be selected at current funding levels. Increased funding to the APRA program would increase the likelihood of developing transformative new technologies for NASA.

H.2.7.2 NSF

NSF funds technology primarily through the Division of Astronomical Sciences Advanced Technology and Instrumentation (ATI) program and the Mid-Scale Innovations Program (MSIP). The Major Research Instrumentation Program (MRI) occasionally funds technology development as well.

The MSIP program was initiated in response to a recommendation from the Astro2010 report. Since 2012, the MSIP program has awarded $146 million to 19 projects. While originally intended to support programs up to $40 million, the largest MSIP award to date has been for $17 million (NANOGrav Physics Frontier Center), with all but three funded programs less than $10 million. Not all of this funding is directed toward technology development, because MSIP also funds construction and operations costs for mid-scale projects. Since 2014, the combined ATI and MRI programs have funded an average of 13 proposals, totaling $7.5 million per year. The award amounts range from $30,000 to $4 million (a large MRI). The wide range of funded award amounts reflects the program's effort to address a range of development activities from simple studies to incremental advancement to deployed instruments on telescopes. However, the average award of $575,000 over 3 years cannot be expected to yield significant technology progress.

The panel suggests expanding the NSF ATI and MSIP programs to enable a significant increase in the number and size of awards to enable substantial technology progress. The panel suggests that an increase of 10 percent per year over the next decade will ensure a healthy program.

H.2.8 NASA Suborbital Program

The suborbital program consistently returns fast-turnaround, cutting-edge science; provides important technology development for future programs; and trains the next-generation of researchers, technologists, and program managers. The suborbital program is critical to maintaining the health of university laboratories capable of carrying out space missions. The suborbital program fills a critical niche by delivering science that is impossible to do from the ground, and does it much more cost effectively than orbital missions.

The suborbital program remains a key part of NASA's portfolio, addressing a wide variety of high-profile scientific problems, developing and testing technology important for future missions and training the next generation of instrumentalists and project leaders. The suborbital portfolio is broken into two components: high-altitude ballooning and sounding rockets. The Enabling Foundation panel received a white paper and met with representatives of both communities who presented the program status.

H.2.8.1 Balloon Program

The introduction of new super-pressure balloons has enabled the exploration of new, more ambitious science missions with significant science returns. The panel suggests that increasing the number of payloads and flights would take better advantage of this capability.

While NASA's high-altitude balloon program has been active for more than 50 years, it has been anything but static. The program has seen a significant evolution in capabilities in response to the user community needs. The Columbia Scientific Balloon Facility (CSBF) is directed by the Balloon Program Office. The program offers a wide array of capabilities ranging from single-day "conventional" flights to Long Duration Balloon (LDB) flights lasting up to 60 days, to the new Super Pressure Balloon (SPB) flights with predicted flight times up to 100 days with extreme altitude stability. They support launch operations from Texas, New Mexico, Sweden, New Zealand, Australia, and Antarctica along with the occasional "remote" launch site. The access to a near-space environment (35–40 km) with such a wide variety of options for duration and sky coverage provides a critical resource to the community. This is demonstrated by the breadth of payload science goals and the hundreds of publications and well over 100 Ph.D.s awarded over the past decade. There is also a small but vibrant program to engage future scientists through "piggyback" instruments on existing payloads.

Maintaining a robust balloon program over the next decade and beyond will require optimization of resources and capabilities within the program:

- There remains a significant technology barrier to new PIs entering the ballooning community. An effort to provide guidance, information, and common hardware and software could greatly facilitate the entry of new researchers into the program. This could include pairing new (or prospective) PIs with more experienced groups to facilitate the transfer of experience and skills. These groups could also be engaged to supply "common" technology such as star trackers and power systems.
- A more formalized outreach program perhaps engaging existing PIs could greatly expand the successful piggyback program to smaller undergraduate institutions throughout the country.
- The SPB capability has advanced significantly over the past decade to the point where 1400 kg payloads are possible. However, the altitudes are limited to 33.5 km, which is too low for many science programs. Developing balloons to achieve higher altitudes and more mass would significantly broaden their utility.
- The number of flights has been dropping over the past several decades. While this has been offset by longer flight durations (total days in the air), a robust program over the next decade will need to expand its user base. This would translate into more funded payloads coupled with more launch opportunities. By encouraging lower cost "conventional" flights with an emphasis on technology development, NASA can increase the size and diversity of the ballooning community.
- Balloon launches place a heavy burden on the CSBF personnel responsible for coordinating all aspects of the program besides the payload itself. Many are required to spend up to 8 months per year supporting launches. This results in high turnover of personnel with highly specialized skill sets. Increasing the number of launches while reducing the burden on the CSBF personnel will be a challenge. It may involve a combination of more trained personnel

with more flights per campaign, such as expanding the number of launches per season in Antarctica and Wanaka with increased infrastructure investment.

H.2.8.2 Sounding Rocket Program

Like the balloon program, the sounding rocket program has been consistently returning science and technology results while training next-generation instrumentation builders over many decades. The rocket program provides crucial access to the space environment with apogees of 250–350 km with 300 seconds above 160 km. This is essentially the only path available to develop and test technologies where even the small amount of residual atmosphere at balloon altitudes is unacceptable. The payloads concentrate in the infrared, ultraviolet, and X-ray. Access to these altitudes comes with relatively short flights and limited payload masses (<220 kg). There are approximately 10 active astrophysics sounding rocket payloads with 4 to 5 flights per year. Compared to the balloon program, the science is not as varied and the direct scientific impact is not as deep. However, unlike balloons, sounding rockets reach space. Therefore, the technology development enabled by the sounding rocket program is more directly traceable to NASA missions, which collectively have significant impact.

The sounding rocket program differs from the balloon program in that the power, telemetry, and pointing systems are provided to the experimental team. This is possible because the sounding rocket capabilities are very well defined. The result is that the barrier for entry by new PIs is, in theory, lower than for the balloon program. However, there has been very little in the way of increased capability for the sounding rocket vehicles (longer flights, larger payloads). While such increased capability could lead to expanded science return, there appears to be limited pressure from the community to do so.

H.2.9 NASA Explorers Programs

NASA's planned cadence provides good balance for the space portfolio.[20] *Pioneers and SmallSats will provide new rapid response opportunities for broad multi-messenger and multi-wavelength observations and technology development.*[21] *The Pioneers program will likely allow teams to complete and fly missions with schedules tailored to the needs of the projects (not limited to the APRA 5-year maximum funding cycle). The panel suggests that a mid-decadal review of the status of the Explorers programs and the impact of the new SmallSat and Pioneers initiatives would be appropriate.*

The NASA Explorers Programs (Table H.1), Medium-Class Explorers (MIDEX), Small Explorers (SMEX), and Missions of Opportunity (MOs), provide consistently excellent scientific returns for a relatively moderate investment and the ability of rapid response to new scientific and technical breakthroughs over a broad range of wavelengths. Explorers enable discoveries with multi-wavelength observations complementing flagship and perhaps future probe missions. The Astrophysics SmallSat program (started in 2018) and Astrophysics Pioneer program (anticipated in 2021) will provide new rapid response opportunities for broad multi-messenger observations and technology development. All of these programs provide opportunities for strategic workforce, scientific, and technical development that ensure the long-term success of NASA's scientific aims. Explorers provide an ideal training ground for the next generation of space experimentalists. There are often university-based, and as such are key to sustaining the university groups that have launched many space science careers.

---

[20] This section was written with substantial input from an interpanel committee of Angela Olinto (Chair), Megan Donahue, Charles Hailey, Bruce Macintosh, Amy Mainzer, Bernie Rauscher, Mark Saunders, and Evgenya Shkolnik.

[21] CubeSats in development: CUTE, SPARCS, BurstCube, SPRITE, and BlackCAT.

TABLE H.1 Status of Astrophysics Explorers, Current and Under Development, and Announcement of Opportunity (AO) Dates

Mission	**AO (Launch) Year**	**Class**		**Mission**	**AO (Launch Plan) Year**	**Class**
NuSTAR	2003 (2012)	SMEX		SPHEREx	2016 (2023)	MIDEX
NICER	2011 (2017)	MO		ARIEL (CASE)	2016 (2028)	MO
TESS	2011 (2018)	MIDEX		ESCAPE or COSI	2019 (2025)	SMEX
IXPE	2014 (2021 plan)	SMEX		Dorado or LEAP	2019 (TBD)	MO
GUSTO	2014 (2021 plan)	MO		(To be selected)	2021 (2028)	MIDEX
XRISM[a]	2008 (2022 plan)	MO		(To be selected)	2021 (TBD)	MO

[a] XRISM is the successor to JAXA/NASA Hitomi (launched in 2016), which failed after a month in orbit. NASA's contribution to Hitomi was selected as an Explorer MO in 2008.

NASA Astrophysics plans an Explorers program cadence of two MIDEX, two SMEX, and four MOs per decade. In addition, NASA plans a cadence of 5 to 10 Pioneers and about 10 SmallSats (i.e., CubeSats <6 U)[22] per decade. The SmallSat program deployed one CubeSat in 2018 (HaloSat) and has five in development for deployment from 2020 to 2024. The Pioneers program will include larger SmallSats (CubeSats >6 U), major balloon payloads, and modest International Space Station attached payloads (with a $20 million FY 2020 cost cap, not including launch). Pioneers are designed to fill in the gap between the Astrophysics Research and Analysis (APRA) program (typically <$10 million) and Explorers MOs (<$35 million for SmallSats). Explorers' most recent cost caps were $250 million (FY 2017) for MIDEX, $145 million for SMEX (FY 2020), $75 million (FY 2020) for associated MO for Small Complete Missions and Partner Missions of Opportunity, and $35 million (FY 2020) for SmallSat MO, all without the launch vehicles. Explorers, Pioneers, and SmallSat programs are openly competed through peer-review processes.

The CubeSat program is the newest addition to NASA's program. While not strictly sub-orbital, there are parallels to both the balloon and sounding rocket programs. CubeSats are volume-limited in their current implementation up to 12 U. The CubeSat includes a science instrument with a bus. The bus can be commercially provided and includes power, communications and pointing. Like the sounding rocket program, this commercially provided infrastructure lowers the barrier to CubeSat entry to new PIs, although the cost is borne by the PI. The management of a CubeSat instrument program could be a

[22] A Unit (U) is a standard volume unit for a CubeSat, a cube 10 cm on a side.

challenge, and it appears to the panel that moving to the sounding rocket model of a NASA-provided BUS would dramatically increase participation.

H.2.9.1 Diversifying the Explorer Program

Explorer missions tend to be led by teams composed mostly of white m from a limited set of universities and NASA centers. Using gender as a marker of diversity, a study[23] of Explorer-class proposals from 2008–2016 finds that the participation by women in the leadership and science teams "is well below the representation of women in astronomy and astrophysics as a whole."

PIs, leadership teams, and instrument builders of selected Explorer missions are less diverse than even the astronomy and astrophysics community. To remove barriers to access, the panel suggests that efforts to diversify Explorers and SmallSat mission leaders and teams be enhanced. The panel suggests that NASA require proposing teams to demonstrate diversity in their teams. A team's diversity includes the diversity of institutions, geographical locations, stages in career, and underrepresented populations.

Given the extraordinarily important role that the Explorers program plays in scientific discoveries, technology demonstration, and in the training of space scientists at all levels, it is vital that Explorers reflect the diversity of the workforce to which the scientific community aspires.

There are substantive barriers to entry to the Explorers program that are unique to the costly and complex development of satellite experiments, and those barriers affect the diversity of teams and scientific ideas. One barrier is the need for considerable engineering resources and expertise that are currently (and deliberately) underfunded and that must be provided by the proposing institutions.

The panel suggests that NASA implement a new program of Concept Maturation Studies (CMS) for future Explorer missions. The new program would select 5 to 10 CMS, for future SMEXs, MIDEXs, and MOs, placing primary emphasis in the evaluation on the proposed science, instrument concept definition and maturation, and team diversity, with less emphasis on management, detailed engineering, and cost. It is important that CMS awards are not a precondition to proposing or being selected for an Explorer call.

The panel suggests that the CMS be solicited and funded separately from the normal Explorer AO process, but with the same cadence as Explorer AOs per decade. The level of effort required for responding to the CMS AO is comparable to an APRA sub-orbital proposal, but with a more extended discussion of science and diversity of the proposing team, the primary criterion in this pre-step 1 program. The reduction in emphasis on technical, management, and cost maturity would properly put the focus on science, diversity, and technical innovation, allowing institutions with fewer resources to compete effectively in the formal AO process. The panel suggests 5–10 fully funded studies, each funded at the $1 million to $2 million level. At the discretion of the PI, a CMS budget could involve resources at NASA centers (such as the Integrated Design Center at Goddard Space Flight Center and TeamX at the Jet Propulsion Laboratory) or other institutions with comparable expertise. This would enable proposals with well-defined science goals to access appropriate engineering support, even if they come from universities that do not have appropriate expertise in house.

The panel suggests that NASA strengthen its support for sustained workshop/workforce development experiences where people with exciting scientific ideas are selected to come to NASA centers and work with engineers to help develop their ideas and their background knowledge in context with current and near-term technologies. The PI Launchpad[24] is an example of such a program. Its goal was to train people interested in developing their first flight mission proposal but have no idea where to start. Continued, ongoing support to develop and run workshops and more intensive internship programs

---

[23] J. Centrella, M. New, and M. Thompson, 2019, "Leadership and Participation in NASA's Explorer-Class Missions," white paper submitted to the Astro2020 decadal survey, https://arxiv.org/abs/1909.10314.

[24] NASA, "PI Launchpad Workshop Content," last update July 29, 2021, https://science.nasa.gov/researchers/pi-launchpad.

of this nature is essential to lowering the barriers to entry into flight projects and to increasing equity, diversity, and inclusion in the community. NASA centers wield a strong influence on who is selected to propose, so opening their doors to new and expanded sources of ideas will expand the breadth and innovation of NASA's science.

H.2.10 NSF Mid-Scale Programs

The mid-scale program has enabled many exciting projects that have had high impact, have led to important scientific discoveries, and have contributed to training students.[25]

In response to Astro2010, the 2012 NSF Portfolio Review,[26] and as one of NSF's Ten Big Ideas, the NSF MPS Directorate and the Division of Astronomical Sciences (AST) have taken several steps to support mid-scale programs. Mid-scale programs are generally understood to be those with total costs between the funding scale of the Major Research Instrumentation (MRI) program (currently $4 million) and the Major Research Equipment and Facilities Construction (MREFC) program (currently $70 million).

An important component of the support for mid-scale programs within both AST and PHY is the Mid-Scale Innovations Program (MSIP), which has been solicited biyearly since 2014. A key review criterion for MSIP awards is the value and benefit to the U.S. astronomical community. MSIP generally has provided $2 million to $10 million total awards to each project, and can support a variety of activities (science project operations, facility, development, or open access capability). Through its first three cycles, it has competitively awarded $114 million to 18 distinct projects that span a diverse range of projects covering almost the entire spectrum available to ground-based astronomy, including 11 projects in radio and 7 in optical astronomy. In addition, NSF awarded $9 million of MSIP funding to the Dark Energy Survey without an open call for proposals, and with only one proposal considered.

A more recent program that supports mid-scale projects is the NSF-wide Mid-Scale Research Infrastructure (MSRI) program, which has been solicited once starting in 2019. MSRI is restricted to fund only design and construction, not operations. The first track of the program (MSRI-1) funded programs up to $20 million and in AST has issued a total of $16.7 million to two astronomy projects, the Event Horizon Telescope (EHT) and development for CMB-S4. The second track of this NSF-wide program, commonly referred to as MSRI-2, calls for implementation proposals with total project costs in the range from $20 million to $70 million. According to NSF, the long-term intent is for MSRI-2 to cover project costs over a range extending up to $100 million.

In addition, NSF has awarded a number of other projects at mid-scale without open solicitations, through special project funding, often to take advantage of internal NSF opportunities for the benefit of AST and to form partnerships with other agencies to leverage their resources. While NSF did not provide requested data on the overall level of NSF investment in such programs, an example of one such enabling partnership is the NN-EXPLORE project to support extreme precision radial velocity (EPRV) work at the 3.5 m WIYN telescope. To date, $8.8 million has been funded from NSF and roughly $15 million to $20 million from NASA. This funding follows a recommendation of Astro2010 to support EPRV, as well as fulfilling the need to support NASA's TESS mission.

In order to assess the health and impact of the mid-scale program, the Enabling Foundations panel of the Astro2020 decadal survey formed a working group that solicited input from other panels, as well gathering info from NSF.

Taken as a whole, the investments made by NSF in mid-scale projects through the MSIP, MSRI-1, and MSRI-2 programs have indeed provided substantial benefits to the U.S. astronomical community

[25] This section is based on a report from an interpanel committee: James Stone (Chair), Michael Blanton, Jenny Greene, David Kieda, Andrea Lommen, Dan Marrone, and David Silva.

[26] NSF, 2012, *Advancing Astronomy in the Coming Decade: Opportunities and* Challenges, National Science Foundation Division of Astronomical Sciences Portfolio Review Committee, https://www.nsf.gov/mps/ast/portfolioreview/reports/ast_portfolio_review_report.pdf.

and fulfilled important and numerous research objectives that cannot or would not be addressed through the AAG, ATI, and MRI grants programs or as part of MREFC projects. Based on an analysis of listed refereed publications available from the websites of the 18 MSIP awardees, they have at least 500 publications published through March 2020. Several MSIP programs provide publicly available astronomical data, and it is likely that these have generated even more publications that are not tracked through the websites. There are several important results enabled by MSIP worth listing:

- Public release of data from the Dark Energy Survey, Zwicky Transient Factory, and the HyperSuprimeCam survey, providing important benefits to the U.S. community not only in their own right, but also as mid-scale forerunners of the Vera Rubin Observatory.
- The Event Horizon Telescope's image of the supermassive black hole at the center of M87 has provided important insights into space-time in the strong gravity regime, the extreme conditions in accreting plasma, and has captured the imagination of the public across the world.
- Direct measurements of the sizes of 300 nearby stars from the Center for High Angular Resolution Astronomy has vastly improved astronomers' ability to test stellar structure and evolution models, as well as image surface features such as flares and starspots.
- With its higher angular resolution, measurements of the cosmic microwave background by the Atacama Cosmology Telescope Polarization (ACTPOL) survey have provided an independent check of the Planck best-fit cosmology. Moreover, POLARBEAR's measurement of the B mode power spectrum is an independent determination of the gravitational lensing signal.

The proposal pressure for the MSIP program has been strong, with a total of 87 proposals submitted in the three cycles, for a funding rate of ~20 percent. The funded projects include new instruments for optical and radio telescopes that come with public access/data, enhancements to national facilities, numerous dedicated cosmological experiments, and open survey programs. Virtually none of these projects fit within the funding envelope of the ATI, MRI, or AAG programs, so the presence of this funding mechanism has been an essential component of astronomical progress in the 2010s.

The panel notes clear signs of strain in this program. Although NSF did not provide any information about the unfunded programs, it is notable that the largest funded projects have budgets that are less than 30 percent of maximum cost specified in the program solicitation ($40 million in the first cycle, $30 million thereafter), a situation that is unimaginable in other AST mechanisms. The MSIP has been unable to fund at least one proposal near its cost cap—CCAT—despite its identification as the top medium-cost ground-based program in Astro2010. There are other examples of programs known to have failed to fit within the MSIP despite strong external reviews or strong subsequent performance, including the Frequency-Agile Solar Radio telescope, which has been ranked highly by multiple decadal surveys in solar and space physics and astronomy and astrophysics, and SDSS-IV, which has yielded a valuable community resource and hundreds of publications without MSIP funding.

The sense of the committee is that the opportunity to productively support mid-scale opportunities is not nearly saturated. Expansion of the MSIP program would be rewarded with proportionally increased scientific productivity of the field.

NSF is the de facto federal steward for the general health and welfare of U.S. ground-based astronomy. Judging from the white papers submitted by the U.S. community to Astro2020, there is great scientific opportunity to be enabled by a strong mid-scale investment program. Moreover, many of these opportunities can or will leverage partnerships with other NSF divisions, federal agencies and private philanthropy. It is also clear that NSF support needed to realize these opportunities is often greater than $20 million per project. All three of these observations (opportunity, partnership potential, NSF investment needed per project) were validated repeatedly over the past decade.

The creation of MSRI-2 in 2019 for projects in the $20 million to $70 million dollar range was a

very welcome development. While it is too early to evaluate the MSRI-2 return-on-impact for astronomical research, MSRI-2 has a healthy oversubscription rate in terms of proposal submission and dollars requested suggesting increased funding for this program would have high scientific impact across several fields, not just astronomy and astrophysics. Initial outcomes have awarded successful MSRI-2 proposals at or near the upper funding bound. Time will tell, but a range of project sizes, enabling projects by more groups, would be a desirable outcome with high scientific impact judging from ideas presented to the Astro2020 panels.

In the MSRI-2 and MREFC programs, projects with widely disparate budgets spanning different divisions and directorates are reviewed together. This could potentially limit mid-scale projects in AST, especially given an apparent internal NSF culture to homogenize total award funding across all divisions. Continuation of a healthy AST-only MSIP program would be very wise if the astronomical sciences community envisions many exciting projects with budgets in the mid-scale range in the coming decade.

How big is “mid-scale”? Following the path of MSRI-1 ($4 million to $19 million) and MSRI-2 ($20 million to $69 million), a new MSRI-3 program with a range of $70 million to $150 million may be warranted. The latter could be split out from the MREFC funding line and indeed could be called MREFC-1 with larger projects going to MREFC-2. Given the larger aspirations of the entire NSF-supported research community, it may be time to separate major projects (under $500 million investment by NSF) from even larger projects ($500 million and above). Last, at the lower end, there is still much scientific opportunity to be realized in astronomy and astrophysics at the $5 million to $20 million range enabled by MSIP. The panel recognizes the importance of having as few siloed funding opportunities as possible; a healthy MSIP program seems well-warranted.

In order to keep costs down, funding for project management has sometime been reduced at the proposal level to such a degree that the success of the effort is endangered. Moreover, while the MSIP program provides funding for operations, the MSRI-1 program does not, highlighting yet one more reason why the MSRI-1 program is not a replacement for MSIP. The importance of recruiting and supporting a strong management team with sufficient resources for operations is often crucial for success; the panel suggests that this aspect of the program be emphasized more in evaluations.

The NSF funding at mid-scale has been a mix of competitive and noncompetitive programs. Competitive programs follow a best practice including openly advertising calls for proposals, multiple considered proposals, and peer review in proposal evaluation. Both the competitive and noncompetitive awards have led to important scientific advances. Circumstances at times will justify future noncompetitive awards, but the panel suggests that the accepted best practice employed by competitive programs be the norm.

Last, the success of the few interagency projects funded at the mid-scale over the past decade suggests that further, and closer, interagency cooperation and funding opportunities could greatly benefit the science that can be supported in the coming decade.

There are more worthwhile projects proposed to the AST MSIP program than can be funded at the current budget levels. The panel suggests enhancing the support of a mid-scale program funded entirely within Astronomical Sciences.

H.2.11 Programmatics of Ground-Based Resources

The disconnect between ground-based observing time and funding for those projects is a significant inefficiency in the current system, a barrier to scientific progress, and a hindrance to the science return on infrastructure investments.

One of the benefits to winning observing time on a NASA instrument is access to a grants program that supports the analysis of that data once acquired and the dissemination of results. But funding is not a benefit of winning time on ground-based facilities run by NSF. The panel identifies this disconnect as a significant inefficiency in the current system. Investigators are often caught in a situation where a grant review panel could be skeptical about a proposed project being awarded the needed

telescope time to complete the project and where a telescope allocation committee could be skeptical about a team having the necessary resources to make the most of the proposed observations. This situation is a net drain on the entire system because projects effectively need to be reviewed twice (by a telescope allocation committee and a grant review committee) and because some teams have data/observing time but insufficient resources to properly collect, analyze, and disseminate any results. This situation is particularly concerning because it has a disproportionate impact on teams with access to fewer institutional and personal resources.

In order to maximize the scientific return of their investment in ground-based observing facilities, the panel suggests that NSF explore funding travel costs and publication costs for U.S.-based teams that have been awarded observing time on public facilities. These costs could be viewed as part of telescope operating costs. Observations from a telescope must be published; otherwise, they have no scientific value. The panel suggests that NSF consider directly funding journals that publish observations from public telescopes. If this were done through directly funding the journals rather than small grants to universities, this would reduce overhead. If this program included a grant component, it could extend the student support programs already implemented for ALMA and NOAO observations.[27,28]

In order to maximize the scientific return of their investment in ground-based observing facilities, the panel suggests that NSF explore funding travel costs and publication costs for U.S.-based teams that have been awarded observing time on public facilities. These costs could be viewed as part of telescope operating costs. Observations from a telescope must be published; otherwise, they have no scientific value. The panel suggests that NSF consider directly funding journals that public observations from public telescopes. If this were done directly through funding the journals rather than small grants to universities, this would reduce overhead. If this program included a grant component, it could extend the student support programs already implemented for ALMA and NOAO observations.[29,30]

H.3 CONCLUSIONS

Over the past decade, the United States has made major investments in astronomical instrumentation, and these instruments are poised to produce data that could transform our understanding of the universe. However, this transformation will require that astronomers, physicists, and computer scientists produce the tools needed to analyze the data, make the laboratory measurements needed to interpret the data, and develop the theory and simulations that enable new paradigms. Investing in the enabling foundation is an investment in the people that will do this transformative science. It is also an investment in the people who will develop the technologies that will enable the observatories, satellites, and instruments of the future.

These investments remove barriers and ensure retention to entry to create a more diverse astrophysics workforce. As advocated in the National Science Board's *Vision 2030*:[31]

> At the post-secondary level, the U.S. must embrace a pathways model to workforce development. Because entry into the STEM workforce is not always via a linear high school–university–workforce path, the U.S. must offer individuals, from skilled technical workers to Ph.D.s, on-

---

[27] National Radio Astronomy Observatory, "Student Observing Support (SOS) Program," last update June 20, 2021, https://science.nrao.edu/opportunities/student-programs/sos.

[28] National Radio Astronomy Observatory, "Financial Support," last update December 16, 2016, http://ast.noao.edu/observing/financial-support.

[29] National Radio Astronomy Observatory, "Student Observing Support (SOS) Program," last update June 20, 2021, https://science.nrao.edu/opportunities/student-programs/sos.

[30] National Radio Astronomy Observatory, "Financial Support," last update December 16, 2016, http://ast.noao.edu/observing/financial-support.

[31] National Science Foundation, 2020, *Vision 2030*, NSF National Science Board, NSB-2020-15, Alexandria, VA, https://www.nsf.gov/nsb/publications/2020/nsb202015.pdf.

> ramps into STEM-capable jobs. The U.S. must also deepen partnerships between educational institutions and the business sector to prepare Americans for the industries of the future and support reskilling and upskilling of incumbent workers so that they can better navigate rapid changes in the world of work.
>
> In order to lead in 2030, the U.S. also must be aggressive about cultivating the fullness of the nation's domestic talent. Although the proportion of Black and Hispanic representation in S&E [science and engineering] jobs rose slightly from 1995 to 2017, these groups remain underrepresented compared to their proportion in the general population. Over the past two decades, the number of women in S&E occupations has doubled. Yet despite comprising over half of the college-educated workforce, as of 2017, women account for just 29 percent of the S&E workforce.

The demographics of astrophysics mirror the rest of science and engineering. By creating more paths to entry into the field, by encouraging retention, and by enhancing programs that attract a diverse range of students, astrophysics will attract a richer range of talents.

I
Report of the Panel on Electromagnetic Observations from Space 1

SUMMARY

The Electromagnetic Observations from Space One Panel (EOS-1) was constituted to examine the state of ultraviolet (UV), optical (O), and infrared (IR) observations from space. This wavelength coverage is from approximately 0.09 to 5 microns. Observations in this wavelength range have been dominated by the Hubble Space Telescope (HST), and the IR wavelengths will be dominated in the near future by the James Webb Space Telescope (JWST). The upcoming Roman WFIRST mission will also make substantial contributions at near-IR wavelengths. The panel is charged with surveying the ability of these current near-term activities, as well as assessing the capabilities of proposed activities, to address the compelling science challenges identified by the science panels convened as part of the decadal survey. The panel is also charged with reviewing the white papers pertinent to UVOIR activities in space. A total of 67 white papers were received by EOS-1. Additional written input included NASA-funded mission study reports from the two flagships and five of the probes considered by this panel. The panel also benefited from the National Academies *Exoplanet Science Strategy* report[1] from 2018. The statement of task is given in Appendix A.[2]

The Astro2020 science panels outlined 24 questions and 6 discovery areas that will define progress for the next decade in astrophysics. Many of these areas need UV-optical-near infrared data. The area of exoplanet research has seen enormous progress toward finding life elsewhere in the galaxy with great strides since the 2010 *New Worlds, New Horizons* survey report. First was the unambiguous detection of an exoplanet in a direct image followed by NASA's Kepler mission census, which reveals that there are large numbers of exoplanets with many more on the way from the Transiting Exoplanet Survey Satellite (TESS). The number of direct images of exoplanets and disks at visible and near-infrared wavelengths has gone from a few in 2010 to a few dozen in 2020, thanks to new ground-based instruments that were deployed over the past decade such as GPI, SPHERE, SCExAO, and LBTI. All of these have advanced science as well as technology. Technological progress has made approaching the Panel on Exoplanets, Astrobiology, and the Solar System (EASS) discovery area of "The Search for Life on Exoplanets" potentially feasible. The panel recognizes the challenges inherent to searching for biosignatures, but progress has made this goal achievable and serves to organize and measure the best of NASA's energies and skills. This goal will capture the imagination of all humankind, and imply technical capabilities that will serve a great majority of the astronomical community beyond the field of exoplanets, and also tie to solar system and earth science.

Other science areas will benefit from proposed missions that are aimed primarily at searching for life, but which provide a compelling suite of capabilities that can address other questions. Astronomy has made rapid progress when panchromatic data are available, and for this panel, of particular importance are the UV spectroscopic capabilities, which will surpass those available on HST, and which will address

[1] National Academies of Sciences, Engineering, and Medicine, 2018, *Exoplanet Science Strategy*, The National Academies Press, Washington, DC.

[2] See Appendix A for the overall Astro2020 statement of task, for the set of panel descriptions that define the panels' tasks, and for additional instructions given to the panels by the steering committee.

many issues identified by the science panels such as observing the circumgalactic medium in emission lines and providing UV imaging and spectra for transient events such as mergers of compact objects.

The EOS-1 Panel considered several implementations of a mission aimed first at detecting biosignatures and with capabilities for expanding the understanding of exosolar systems in general and with capabilities for enabling broad ranging observations, especially significant at UV wavelengths. The panel is suggesting further study and technology development that could lead to a mission with the light gathering power of at least a 6-meter primary mirror, and equipped with a light suppression system capable of achieving the goal of the 10^{-10} contrast needed for detection of Earth-like planets in the habitable zone of solar-type stars. The mission will also need focal plane instrumentation to acquire images and spectra over the range of 100 nm to 2 microns with parameters similar to cameras and spectrometers proposed for the Large Optical UV Infrared Telescope (LUVOIR) and the Habitable Exoplanet Observatory (HabEx). These instruments would include moderately wide-field imaging at UV, optical, and near-IR wavelengths as well as multi-object spectroscopy over a similar wavelength range (see Table I.5, later). The panel is not suggesting a named mission now because it is premature to do so this far ahead of when actual development could start, a situation caused by the fact that two astrophysics flagships are still being completed, as well as budgetary considerations. By not naming a particular mission configuration now, the panel is stating that more work is needed to ensure that a future mission's budget envelope is well-constrained and that the mission will achieve its primary science goals given the uncertainties in parameters such as η_{Earth}. These points are discussed in greater detail in Section I.3.

The panel reached its conclusion for a future flagship mission based on the following observations, which are described more fully later in this report. The panel viewed it as essential that a flagship-class mission be capable of achieving a compelling result for the first time while also being capable of supporting a broad range of other science investigations. First, both the LUVOIR and HabEx teams presented a convincing case that astronomers are positioned to make a serious attempt at searching for biosignatures on exoearth candidates, which is the compelling result projected for these missions. This goal aligns with one of the principal recommendations of the *Exoplanet Science Strategy* report:

> **Recommendation: The National Aeronautics and Space Administration (NASA) should lead a large strategic direct imaging mission capable of measuring the reflected-light spectra of temperate terrestrial planets orbiting Sun-like stars.**

The EASS Panel states that searching for evidence of life "is only possible with a large high-contrast direct-imaging space telescope."[3] Neither JWST, Roman WFIRST, nor the Extremely Large Telescopes (ELTs) are predicted to have coronagraphs with adequate contrast, whereas small missions will have inadequate light-gathering capability for the needed spectroscopy of an exoearth. Second, any of the mission configurations that the panel evaluated that would be capable of finding more than one or two candidate exoearths require too much funding in the peak development years. Even the smallest mission considered in detail, HabEx 3.2S, requires a substantial increase in the budget allocation for new missions or else the mission would use almost all of the Astrophysics budget in its peak years according to the Technical, Risk, and Cost Evaluation (TRACE) analysis (see Appendix O). These budget concerns are exacerbated by the panel's finding that a probe-class mission line with two launches per decade might also be added to the Astrophysics portfolio. Third, substantial technology development is still needed despite recent progress in starlight suppression techniques and ultra-stable telescopes with at least 5 years and several hundreds of millions of dollars needed. The panel recognizes that for this level of technology development, it would be usual to choose a mission architecture to focus the technology development, but given the budgets and time scales needed, choosing a specific mission now is premature. But equally, the panel would like to ensure that substantive progress is made toward an eventual choice of a flagship capable of detecting exoearths and biosignatures, so the panel has considered how to achieve getting

[3] Appendix E, p. 18.

technologies to the point where a mission could start as early as the late 2020s and preferably launch no more than a decade later.

Observation: Considerable progress has been made in improving starlight suppression performance, which suggests that the desired 10-10 contrast ratio needed for direct detection of earth-like exoplanets around Sun-like stars is achievable with adequate resources.

Observation: Progress in defining biomarkers as outlined in *Exoplanet Science Strategy* (2018) suggests that 0.1 to 2 microns is a rich wavelength regime, including the UV ozone feature as a robust indicator of oxygen in an exoplanet atmosphere (see Figure I.1).

The panel also reviewed 11 probe-class missions. Two versions of the CETUS probe concept were presented to the panel. The science goals and observational capabilities for the two versions are almost identical with implementation being the major difference. In the remainder of this report, the two versions are treated as the same suite of observational capabilities. Collectively, these probe missions present a suite of capabilities that would address a number of the science questions identified by the science panels. While probes cannot fulfill all of the observational needs from space for astrophysics, they could provide very valuable data, such as multi-wavelength observations for time-critical observations that will be difficult to provide from flagships only. Probes have the potential to broaden the suite of capabilities available to astronomers, and to respond to changing science priorities more quickly than flagship missions can. The panel prefers open competition for probe launches.

Observation: A variety of probe-class missions would be capable of delivering some but not all of the high-priority science identified in the Science Panel reports (see Error! Not **a valid bookmark self-reference.** presents a science panel mapping for the probe-class mission considered by the panel. The entries in the table refer to the focal plane instrument providing the relevant capability. A white background indicates a major contribution to the question, whereas gray indicates a lesser but still significant contribution. Gray has a similar meaning as for the flagships: some capability is provided with similar wavelengths and spectral resolutions as suggested by the science panels.

The COEP Panel highlighted the need for multi-wavelength observations on a variety of time scales. That panel cites the need for rapid follow-up of events at UV wavelengths which will be impossible after HST ceases operation. Panchromatic observing capabilities needed in the coming decade were highlighted in WP Megeath. This need touches on a number of science themes identified by the science panels. The success of the great observatories (Spitzer, Hubble, Chandra, and Compton) is a strong motivator for providing panchromatic capabilities to the astronomical community in the coming decade. Some probe-class missions such as TAP seek to provide panchromatic capabilities in a single facility to address specific science themes. Overall, the panel determined that there is strong scientific motivation for facilities, especially space-based, to provide the broadest wavelength coverage possible.

TABLE I.2, later).

I.1 PANEL INPUTS AND CONSIDERATIONS

The panel met face-to-face two times, remotely once, and conducted eight teleconferences to examine the state of UVOIR activities from space. The face-to-face and remote meetings concentrated on presentations from representatives of the two flagship missions that fall under the panel's purview, LUVOIR and HabEx, and on presentations from representatives of all of the Probe-class (~$1 billion missions) that fall in the UVOIR wavelength range. For the flagship missions, the panel sent questions to the teams both before and after presentations, although "export control" markings hampered evaluation of some responses. Resources to subject missions to the TRACE process were limited, so the panel chose to focus primarily on LUVOIR-B and on two versions of HabEx, 4H and 3.2S, as a balance between cost and exploring a range of implementation options. Some of the probe missions received some limited study funding from NASA, while others were submitted to the decadal survey solely as white papers. The panel considered probe-class missions that received NASA funding, including CETUS, CDIM, TAP, Starshade Rendezvous, and Earthfinder (proof of concept study only). Other probe-class missions considered that did not receive NASA study funding include EXO-C, OOO, an alternate formulation of CETUS, ANUBIS, ATLAS, and Nautilus.

The panel also heard from invited speakers on some of the techniques germane to missions under consideration. Talks on how starshades and coronagraphs work were included in the first meeting. The first meeting also included presentations on the status of technologies for these two starlight suppression techniques. Because of its significance for predicting the exoearth yield from the flagship missions, the panel heard a presentation on the current state of knowledge of η_{Earth} at the second meeting. An additional talk on the Roman WFIRST coronagraph was included in the panel's third meeting. The panel also read 67 white papers submitted by the community.

The Aerospace Corporation briefed the panel on its Technical Risk and Cost Evaluation results for LUVOIR B and two versions of HabEx at the panel's third meeting. The panel also had access to the Large Mission Concept Independent Assessment Team report on LUVOIR and HabEx, and the Probe Cost Assessment Team report on the NASA-funded probe studies.

The panel also assessed the current near-term space capabilities in the UV-optical-IR wavelength range to see how the flagship and probe missions presented to the panel fit into the current observational opportunities. The range covered by the James Webb Space Telescope (0.6 to 28 microns) will be well-supported by the capable imagers and spectrometers provided on that mission. Study of exoplanet atmospheres via transmission spectroscopy observed in a transit will be the prime exoplanet observational mode on JWST. Roman WFIRST will include a demonstration coronagraph and will provide near-IR surveys over broad areas on the sky, and a range of small missions are enhancing many aspects of exoplanet investigations such as TESS, ARIEL, PLATO, and CHEOPS.

I.1.1 Major NASA Operating Missions

The Hubble Space Telescope (HST) remains one of the world's premier astronomical observatories 30 years after it was placed in low Earth orbit by the space shuttle. Its suite of imaging and spectroscopic instrumentation, restored to full functionality in the final HST Servicing Mission in 2009, covers the UV to near IR (0.1—1.7 μm). HST's scientific accomplishments are legion, and the observatory remains in great demand by astronomers across the world. It is the only UV spectroscopic capability for the foreseeable future. While HST is no longer serviceable with current capabilities, its orbit is stable against reentry into the 2030s, although a gyro failure could limit its operational lifetime.

The Stratospheric Observatory for Infrared Astronomy (SOFIA) is a telescope with an effective diameter of 2.5 meters, carried aboard a Boeing 747-SP aircraft. The observing altitudes for SOFIA are

between 37,000 and 45,000 feet, above 99 percent of the water vapor in Earth's atmosphere. The present instrument suite provides coverage from 0.3–612 μm. New instrumentation is currently in development. The current instrumentation is complementary to the future capabilities of JWST, providing imaging and spectroscopy (R ~ 100 and R >10,000) in the mid-IR. Other instruments cover 30–600 μm with imaging, polarimetry, and high-resolution spectroscopy. SOFIA is made possible through a partnership between NASA and the German Aerospace Center (DLR). The observatory's mobility allows researchers to observe from almost anywhere in the world. SOFIA has conducted regular observing campaigns in the Southern Hemisphere.

The Neil Gehrels Swift Observatory was launched in 2004 to study gamma-ray bursts (GRBs) with a range of gamma-ray and X-ray instrumentation. The mission has been a huge scientific success. Of particular interest to the EOS-1 Panel is the UV-optical telescope that is part of Swift's instrumentation suite. The UVOT (0.17–0.6 μm) takes images and grism spectra of GRB afterglows during pointed follow-up observations. The images are used for 0.5 arcsecond position localizations and following the temporal evolution of the UV-optical afterglow. Spectra can be taken for the brightest UV-optical afterglows, which can then be used to determine the redshift via the observed wavelength of the Lyman-alpha cut-off. The UVOT has made significant contributions to the study of a range of transient phenomena, not just GRBs.

NASA's Transiting Exoplanet Survey Satellite (TESS) is designed to search for exoplanets using the transit method of 200,000 nearby stars over 85 percent of the sky, an area 400 times larger than that covered by the Kepler mission. It was launched in 2018 and is expected to find more than 14,000 transiting exoplanets, compared to about 3800 exoplanets known when it launched. As of March 2021, TESS has identified more than 2645 candidate exoplanets, of which 122 have been confirmed so far. In January 2020, NASA announced the discovery by TESS of the first Earth-size planet in its star's habitable zone. TESS will provide prime targets for further characterization by the JWST, as well as other large ground-based and space-based telescopes.

I.1.2. International Operating Missions

ESA's Characterizing Exoplanet Satellite (CHEOPS) is the first mission dedicated to studying bright, nearby stars that are already known to host exoplanets, to make high-precision observations of the exoplanet's size as it passes in front of its host star. CHEOPS was launched in December 2019. It is focused on planets in the super-Earth to Neptune size range, with its data enabling the bulk density of the planets to be derived.

ESA's Gaia is an astrometric mission to measure the positions and velocities of ~1 billion stars in the Milky Way galaxy. These measurements are being used to construct a three-dimensional map of the galaxy. The observational phase of the mission is complete, but data analysis is still in process, with the full DR3 release expected in 2022. The mission is having both broad science impacts and an impact on operations of future missions with exquisite star positions that enable much better pointing than previously achievable.

I.1.3 Approved Missions in Development

The James Webb Space Telescope (JWST) is a NASA Flagship mission, in partnership with ESA and CSA, in development for launch in 2021. JWST will provide significantly improved infrared angular resolution (0.031 arcsec at 2 μm) and sensitivity over HST and the now retired Spitzer. It is designed to enable a broad range of investigations across astrophysics, including finding and studying the first galaxies in the early universe. JWST will provide major capabilities for studying exoplanets and planet formation across its 0.6–28 μm wavelength range. All four scientific instruments have observing modes

designed for exoplanet transits. Three instruments have coronagraphic or aperture mask imaging to obtain direct images of exoplanets. JWST will orbit the Sun at the Sun-Earth L2 point.

JWST is a new type of space telescope design. To achieve the cryogenic temperatures (<50,000) necessary for low-background IR observations, the telescope and scientific instruments are open to deep space and cool passively, and remain cold, because they are shielded from the Sun's and Earth's heat by a large sunshield. The sunshield and the telescope are folded for launch and are deployed in space. The deployable sunshield, telescope optics, and large composite structures are among the key technologies being demonstrated by JWST that are relevant for future mission designs.

Nancy G. Roman WFIRST is a NASA mission to survey wide swaths of the sky at near-infrared wavelengths to address fundamental questions about the nature of dark energy, to provide a statistical basis for understanding exoplanetary system architectures via microlensing that is free of the biases of transit studies, to demonstrate a more capable coronagraph than ever used previously in space (CGI), and to provide wide-area near-infrared imaging for guest observer programs. Wide-field imaging will be performed from 0.5 to 2 μm with a spatial resolution of 0.11 arcsec. All of these programs are as vital as they were when the mission was ranked highly in *New Worlds, New Horizons* and when the coronagraph was added. The panel finds the mission compelling, and the CGI is a useful technology demonstration that will test deformable mirrors in the space environment and also closed-loop wavefront control. The anticipated launch date is in late 2025.

Euclid, named after the ancient Greek mathematician, is a visible to near-IR mission currently under development by the ESA for launch in 2022. The objective of the Euclid mission is to better understand dark energy and dark matter by accurately measuring the acceleration of the universe using gravitational lensing, baryon acoustic oscillations, and measurement of galactic distances by spectroscopy. Euclid will measure the shapes of galaxies at varying distances and investigate the relationship between distance and redshift out to $z \sim 2$. The link between galactic shapes and their corresponding redshift may reveal how dark energy is related to the acceleration of the universe. Euclid employs a 1.2 m telescope as compared to Roman's 2.4 m telescope, and includes a slitless grism capability. Euclid does not have a coronagraph.

Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) will perform an optical to near-IR all-sky survey to measure the spectra of approximately 450 million galaxies. SPHEREx is a NASA MIDEX-class mission planned for launch in 2023. SPHEREx will use a spectrophotometer for its all-sky survey that will record images in 96 wavelength bands from 0.75 to 5.0 μm. It has a single instrument with a single observing mode and no moving parts to map the entire sky four times during its nominal 25-month mission. The key technology is a linear variable filter that shifts the wavelength of the imaging bandpass. SPHEREx will measure galaxy redshifts, categorize galaxies, and fit measured spectra to a library of galaxy templates. SPHEREx will probe signals from the intra-halo light and from the epoch of reionization. SPHEREx will also contribute disk observations and data on the molecular content of stellar nurseries.

Planetary Transits and Oscillations of Stars (PLATO) is an ESA mission designed to detect Earth-like planets in the habitable zones of solar-type stars. By combining space-based visible light photometric transit measurements with ground-based radial velocity measurements, an assessment of the bulk properties of the detected planets will be possible that in turn will provide an indication of which planets might be habitable. PLATO, although superficially similar to TESS, differs in that it is not an all-sky survey like TESS but rather will point at a suite of FGK stars for long periods with the goal of detecting small planets with long periods. The mission is scheduled for launch in 2026.

Atmospheric Remote-Sensing Infrared Exoplanet Large-Survey (ARIEL) is an ESA mission in development for a 2029 launch. It has a 1.2 m × 0.7 m off-axis primary mirror equipped with infrared spectrometers covering 1.2 to 7.8 microns and visible light photometry. The telescope is cooled to 55 K and the focal plane detectors are cooled to ~42 K. The goal is to survey 1000 transiting planets orbiting F to M stars. The survey will provide a statistical sample of exoplanetary atmospheres that can be used to address questions such as how the stellar environment affects exoplanet atmospheres, whether

atmospheric compositions shed light on possible planetary migration, how atmospheres may evolve over time, and many related issues.

Observation: The current suite of missions including those soon to be launched do not include any replacement for the aging UV spectroscopic capabilities available with HST nor do they include any far UV capabilities.

I.1.4 Technology Progress

The 2010 decade has seen rapid and significant advances in many technology areas relevant for UVOIR astronomy, especially in exoplanet detection and characterization. A segmented telescope design has been realized with JWST. Other work has examined how to achieve even greater wavefront stability. Testbed coronagraphs are within a factor of 10 of the desired contrast, and adaptive wavefront control systems capable of removing even secondary mirror strut artifacts have been developed. Post-processing techniques have been demonstrated on HST data to reveal sources previously missed. UV coatings for dramatically improved throughput have been tested. A number of white papers highlighted these technology developments that are important for the missions that the panel considered:

- *Telescopes:* Development of technologies/concepts for ultra-stable telescopes (e.g., ULTRA study; WP[4] Coyle; WP East; WP Feinberg; WP Nordt; WP Wells); active mirror technologies (WP Lawrence).
- *Starlight suppression:* Starshades (WP Short). Subscale demonstration achieved 1.2×10^{-10} contrast; stowage and deployment design developed and demonstrated on sub-scale prototypes; coronagraphs (WP Shaklan, Mazoyer). Broadband lab demonstrations include 3.8×10^{-10} contrast (monolithic aperture, DST/HCIT); $\sim 1 \times 10^{-9}$ static and $\sim 1 \times 10^{-8}$ dynamic contrasts (WFIRST aperture). In addition, coronagraph designs have significantly improved in performance, especially for obstructed apertures, and are continuing to improve every year. See also Table I.7, later in this report.
- *Adaptive wavefront control:* (e.g., WP Pueyo; WP Kasdin) MEMS deformable mirrors (DMs) successfully operated in vacuum in the lab and on a stratospheric balloon (PICTURE); number of actuators has increased from typically ~1K to ~2K; algorithms advanced to where DMs can remove struts and segmentation of apertures, suppress binary stars, as well as mitigate telescope instability. The latter is especially important and is a lever that has not yet been fully appreciated or utilized (WP Pueyo; Crooke et al.), and has the potential to bring stability requirements of LUVOIR and HabEx within range of current JWST segment drift requirements.
- *Post-processing methods:* Ground-based post-processing of directly imaged planets has matured significantly, with powerful methods such as KLIP (Soummer et al., 2012) now being standard.
- *UV coatings:* (WP Sheikh) Can improve telescope throughput and enable a smaller telescope to achieve results that would have required much larger telescope in the past.
- *Photonic and related devices:* (WP Jovanovic; WP Van Buren) May enable breakthroughs in the size and robustness of focal plane instrumentation.

I.1.5 Science Panel Inputs

A crucial component in guiding the formulation of a program for the future is the identification of the most compelling science questions to be addressed in the next decade. The high-priority questions and

[4] WP = white paper.

discovery areas identified by the science panels and corresponding observational capabilities have been mapped to the capabilities proposed for the flagship missions under this panel's purview. Table I.1 presents the mapping with the questions identified by panel as shown after the table. The entries in the table show which proposed instrument (see Table I.5, later) on the flagship would be used to address these questions. The wide-field imagers and multi-object spectrometers on these missions are very similar which is why the missions address similar science questions. To a large extent, missions make progress on the question in proportion to their light-gathering capability. Note that gray entries indicate that the flagship does not satisfy all the requirements needed for the science panel questions, but does provide some capability to address the questions posed, typically by virtue of covering some of the relevant wavelengths at similar spectral resolutions. The table indicates that the flagships considered have a wide scientific reach that goes beyond the exoearth search. The science panels each provided four question areas, and missing entries such as COS-1 indicate that the EOS-1 Panel missions provide no relevant capability in that area. Note that although LUVOIR-B and HabEx missions can address similar science areas, the larger size of LUVOIR-B enables deeper studies with higher quality data which come closer to answering the questions posed by the science panels. The HabEx missions are too small to address some of the science areas effectively, such as those addressing z~7 UV luminosity functions which is part of GAL-1.

Another way to judge the breadth of the science achievable with a flagship mission comes from Table I.2, taken from the LUVOIR Final Report, which lists "Signature Science Cases" indicative of the sweep of science that can be addressed with such a flagship. Science cases 1–3 are associated with the grand goal of Earth-like planet detection, while cases 4 and 5 address other exoplanet science, attainable even if the contrast goal were not met. These cases map directly onto the questions posed by the EASS Panel. Case 6 contributes to other questions posed by the EASS Panel about the relationship of the solar system's architecture to other systems' architectures. Cases 7, 10, 11, and 12 are related to Galaxy Panel questions. Case 8 contributes to work on the Cosmology Panel's question 2. Case 9 is related to GAL-1, but the GAL question is pointed at higher redshifts ($7 < z < 9$), while Case 9 is aiming at sources up to $z \sim 7$. The EOS-1 wavelength range is crucial for studying galaxy evolution as reflected in the capabilities of JWST's instruments. Once JWST begins returning data, this field will no doubt be completely revolutionized (Förster Schreiber and Wuyts, 2020; Tacconi et al., 2020; Volonteri et al., 2017). The Panel on Galaxies has indicated that wide-field and very wide field spectroscopy at 0.32 to 5 microns will be essential for addressing questions such as measuring the characteristics of the first stars, galaxies, and black holes. This wavelength regime is also important for taking a census of SMBH growth and determining the threshold for galaxy formation with LUVOIR and HabEx covering wavelengths up to 2 microns. The Panel on Galaxies also points out the need for measurements at 0.09 to 0.32 micron to connect low-redshift galaxies to high-redshift galaxies. The discovery area for the Panel on Galaxies is mapping the circumgalactic and intergalactic media in emission, which also requires UV spectroscopy at these wavelengths.

TABLE I.1 Mapping of Flagships to Science Panel Questions and Discovery Areas

Question	LUVOIR-B	HabEx-4H	HabEx-3.2S
COS-2	LUMOS	UVS	UVS
GAL-1	LUMOS	UVS, HWC	UVS,HWC
GAL-2	LUMOS	UVS	UVS
GAL-3	LUMOS	UVS	UVS
GAL-4	LUMOS, HDI	UVS, HWC	UVS,HWC
ISP-1	LUMOS, HDI	UVS	UVS
ISP-3	LUMOS,HDI	UVS, HWC	UVS,HWC

ISP-4	LUMOS, HDI	UVS, SSI	UVS, SSI
SSP-1	LUMOS	UVS	UVS
SSP-2	LUMOS	UVS	UVS
SSP-3	LUMOS	UVS	UVS
SSP-4	LUMOS	UVS	UVS
COEP-2	LUMOS, HDI	UVS, HWC	UVS, HWC
EASS-1	LUMOS, HDI	SSI, UVS	SSI, UVS
EASS-2	LUMOS, HDI	SSI, UVS	SSI, UVS
EASS-3	LUMOS, HDI	SSI, UVS	SSI, UVS
EASS-4	LUMOS, HDI	SSI, UVS	SSI, UVS
GAL DA	LUMOS	HWC	HWC
ISP DA	LUMOS, HDI	HWC	HWC
SSP DA	LUMOS	UVS	UVS
COEP DA	LUMOS, HDI	UVS, HWC	UVS, HWC
EASS DA	LUMOS, HDI	SSI, UVS	SSI, UVS

NOTE: LUMOS—LUVOIR Ultraviolet Multi-Object Spectrograph; HDI—High-Definition Imager; UVS—Ultraviolet Spectrograph; HWC—HabEx Workhorse Camera; SSI—Starshade Instrument.
COS—Cosmology; GAL—Galaxies; ISP—Interstellar Medium and Star and Planet Formation; COEP—Compact Objects and Energetic Phenomena; SSP—Stars, Sun, and Stellar Populations.
EASS—Exoplanets, Astrobiology, and the Solar System.

TABLE I.1 Proposed LUVOIR Signature Science Cases

Signature Science Case
1 – Finding habitable planet candidates
2 – Searching for biosignatures and confirming habitability
3 – The search for habitable worlds in the solar system
4 – Comparative atmospheres
5 – The formation of planetary systems
6 – Small bodies in the solar system
7 – Connecting the smallest scales across cosmic time
8 – Constraining dark matter using high precision astrometry
9 – Tracing ionizing light over cosmic time
10 – The cycles of galactic matter
11 – The multiscale assembly of galaxies
12 – Stars as the engines of galactic feeback

Error! Not a valid bookmark self-reference. presents a science panel mapping for the probe-class mission considered by the panel. The entries in the table refer to the focal plane instrument providing the relevant capability. A white background indicates a major contribution to the question, whereas gray indicates a lesser but still significant contribution. Gray has a similar meaning as for the flagships: some capability is provided with similar wavelengths and spectral resolutions as suggested by the science panels.

The COEP Panel highlighted the need for multi-wavelength observations on a variety of time scales. That panel cites the need for rapid follow-up of events at UV wavelengths which will be impossible after HST ceases operation. Panchromatic observing capabilities needed in the coming decade were highlighted in WP Megeath. This need touches on a number of science themes identified by the science panels. The success of the great observatories (Spitzer, Hubble, Chandra, and Compton) is a strong motivator for providing panchromatic capabilities to the astronomical community in the coming decade. Some probe-class missions such as TAP seek to provide panchromatic capabilities in a single facility to address specific science themes. Overall, the panel determined that there is strong scientific motivation for facilities, especially space-based, to provide the broadest wavelength coverage possible.

TABLE I.2 Mapping of Probe-Class Missions to Science Panel Questions and Discovery Areas

Question	TAP	Starshade Rendezvous[a]	Exo-C	OOO[b]	CETUS	ANUBIS	Earthfinder	CDIM	Nautilus	ATLAS
COS-2	N/A	N/A	N/A	N/A	PSS	FUV Spec	Spec	N/A	N/A	N/A
GAL-1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Spec	N/A	Spec
GAL-2	N/A	N/A	N/A	N/A	PSS	FUV Spec	Spec	Spec	N/A	N/A
GAL-3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Spec
GAL-4	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Spec	N/A	Spec
ISP-1	N/A	N/A	N/A	N/A	N/A	Camera	Spec	N/A	N/A	N/A
ISP-3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Spec
SSP-1	N/A	N/A	N/A	N/A	MOS, PSS	FUV Spec	Spec	Spec	NAVIIS	Spec
SSP-2	N/A	N/A	N/A	N/A	MOS, PSS	FUV Spec	Spec	Spec	NAVIIS	Spec
SSP-3	N/A	N/A	N/A	N/A	MOS, PSS	FUV Spec	Spec	Spec	NAVIIS	Spec
SSP-4	N/A	N/A	N/A	N/A	MOS, PSS	FUV Spec	Spec	Spec	NAVIIS	Spec
COEP-2	IRT	N/A	N/A	N/A	Camera, MOS, PSS	Camera, FUV Spec	Spec	Spec	NAVIIS	Spec
EASS-1	IRT	CGI	Imager, IFS	3-Band Phot	N/A	N/A	N/A	N/A	NAVIIS	Spec
EASS-2	IRT	CGI	Imager, IFS	3-Band Phot	PSS	FUV Spec	Spec	N/A	NAVIIS	Spec
EASS-3	IRT	CGI	Imager, IFS	3-Band Phot	N/A	FUV Spec	Spec	N/A	NAVIIS	Spec
EASS-4	IRT	CGI	Imager, IFS	3-Band Phot	N/A	FUV Spec	Spec	N/A	NAVIIS	Spec
GAL DA	N/A	N/A	N/A	N/A	MOS	FUV Spec	N/A	N/A	N/A	N/A
SSP DA	N/A	N/A	N/A	N/A	MOS, PSS	FUV Spec	Spec	Spec	NAVIIS	Spec
COEP DA	IRT	N/A	N/A	N/A	Camera, MOS, PSS	Camera, FUV Spec	Spec	Spec	NAVIIS	Spec
EAS DA	IRT	CGI	Imager, IFS	3-Band Phot	N/A	N/A	N/A	N/A	N/A	

[a] Used in conjunction with WFIRST; N/A = no contribution.
[b] Used with CETUS or CASTOR (Canadian mission). Instruments: TAP: IRT (Optical Infrared Telescope); Starshade Rendevous: CGI (Coronagraph Instrument); EXO-C: IFS (Integral Field Spectrograph); OOO: 3-Band Phot (3-Band Photometer); CETUS: MOS (Ultraviolet Multiobject Spectrometer), PSS(Point Source Spectrometer); ANUBIS: FUV Spec (Far Ultraviolet Spectrometer); Earthfinder: Spec (Spectrometer); CDIM: Spec (Spectrometer); Nautilus: NAVIIS (Nautilus Visual-Near-Infrared Imager and Spectrograph); ATLAS: Spec (Spectrometer).

I.2 BUDGETARY CONSIDERATIONS

The large mission (flagship) category was the most difficult challenge presented to the panel, because all of the options under consideration proved to be significantly higher cost than even the most

optimistic NASA budget can accommodate, particularly if a probe line is to be included in future budgets. The TRACE analyses revealed that the mission costs were likely underestimated by their project teams by nearly 50 percent, largely owing to a combination of TRACE including monetized risks and longer development times that would be driven by budgetary constraints. The largest of these missions (mirror ≥6 m) do an incredible job of searching for Earth-like habitable planets (see detection rates in Table I.4), but require a considerable investment to advance the enabling technologies to a readiness level that will permit them to achieve these results. Given that these technologies are not yet mature (the LUVOIR and HabEx final study reports tabulate TRLs that range from 3 to 5), the current cost and schedule estimates assume that all technologies come to fruition as planned, and thus, the cost and schedule estimates are still very immature. Unfortunately, based on the panel's and Aerospace's analyses, the panel believes that large-aperture telescopes (even as small as 4 m) as presented to the panel are not affordable in a reasonable time frame with the currently forecast NASA Astrophysics budgets. They either take too long to develop or consume too much of the budget, making the Astrophysics Program unbalanced. JWST is a case in point—it will have taken more than 20 years to develop and launch. For current flagship design approaches to be affordable and including a probe line would require that the Astrophysics budget at least double to $3 billion per year. However, the panel believes that a flagship mission, designed to observe habitable exoplanets while providing UV, optical, and near-IR imaging and spectroscopy could be pursued once the following criteria are met:

- All enabling technologies have been advanced to a TRL of 6 prior to authorization to proceed with phase A, which will force earlier investment in technology and a change in NASA policy.
- Sufficient budget is planned to be available to permit completion of the mission development and launch within 10 years of Key Decision Point B, which might require difficult decisions to be made because funding is not formally in place until KDP C.
- The mission has been designed to have a mission lifetime no less than the time to develop the mission as measured from KDP B. (This criterion is not meant to drive mission reliability but rather to drive getting missions developed more expeditiously.)
- In addition to the three suggestions above, other improvements in the management of Astrophysics flagships can be found in Bitten et al. (2019);[5] WP Tumlinson; WP Hylan; WP Crooke.

The panel also observes that NASA could consider pursuing different design approaches in the future for very large aperture telescopes, because the Astrophysics budget may not increase as rapidly as telescope costs. If very large aperture telescopes are required in the future, different design approaches could be considered, including assembly in space, and servicing and modularity that would allow telescopes to evolve, including adding aperture, upgrading capabilities, and extending the life of the telescope. Existing capabilities for assembly of high-performance optical systems in space are nonexistent and will require substantial development if this type of construction is to be realized.

I.3 FLAGSHIPS

Flagships provide capabilities that cannot be achieved at smaller scales with sensitivity and angular resolution typically being the drivers for large telescopes. Flagships include a range of instrumentation, whereas the requirements are set by the most scientifically compelling goals. The instrumentation enables a broad range of other science goals to be addressed, so large missions have

[5] R.E. Bitten, S.A. Shinn, and D.L. Emmons, 2019, "Challenges and Potential Solutions to Develop and Fund NASA Flagship Missions," pp. 1–13, *2019 IEEE Aerospace Conference*, Big Sky, MT, March 2–9, 2019, doi: 10.1109/AERO.2019.8741920.

broad support within the astronomical community. Flagships provide large numbers of astronomers with their first exposure to space data and space projects.

I.3.1 Flagship Science Capabilities

The EOS-1 Panel was presented with two projects that would lead to missions capable of detecting biosignatures on earth-like planets, a compelling goal on many levels. Figure I.1, which appears in both the LUVOIR[6] and HabEx[7] final reports, shows the richness of an exoearth spectrum in the 0.2 to 2 micron region, including the very strong ozone potential biosignature at 0.25 micron. This spectrum assumes that one is observing reflected light, which is possible with direct imaging using a high-performance starlight suppression system. Small-scale height features such as these absorptions would not be observable in a transmission spectrum from a transit observation for an exoearth around a Sun-like star. Earth-like planets around Sun-like stars transit very rarely (from geometry), infrequently (once per year), and at such a shallow transit depth that their atmospheric features are essentially not characterizable by plausible missions.[8]

The LUVOIR team developed two concepts for consideration: LUVOIR-A with a primary mirror diameter of 15 m, and LUVOIR-B with an off-axis primary mirror with a diameter of 8 m. Both concepts rely on coronagraphs for starlight suppression. The LUVOIR team's preferred configuration is the 15 m version. The HabEx team developed nine separate concepts ranging in size from 2.4 m to 4 m and with starlight suppression using either a coronagraph or a starshade. The HabEx team's preferred configuration (4H) uses a 4 m mirror and both coronagraphy and a starshade. The panel examined LUVOIR-B, HabEx 4H, and HabEx 3.2S in detail. The choice of LUVOIR-B over LUVOIR-A was based largely on cost considerations because LUVOIR-A would take too long to build and test without an implausibly large increase in the NASA Astrophysics budget as judged by the team's values and confirmed by TRACE. The panel choose two HabEx configurations, as they provided a comparison between a mission with two starlight suppression techniques (4H) and one with only a starshade (3.2S), and also provided a comparison between a monolithic primary (4H) and a segmented primary (3.2S).

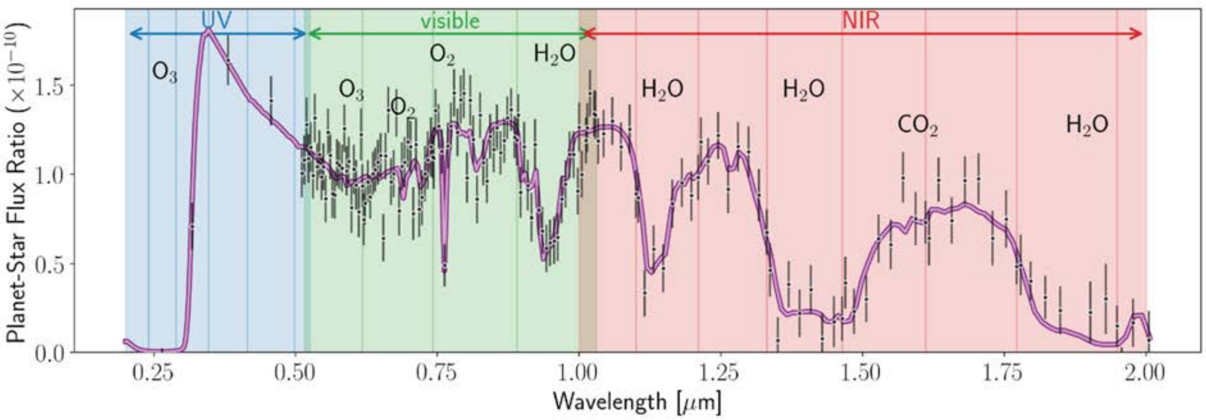


FIGURE I.1 Simulated UV-NIR exoearth spectrum that highlights absorption from several key molecules for biosignature detection such as ozone, molecular oxygen, water, and carbon dioxide. SOURCE: LUVOIR and HabEx final reports. Courtesy of J. Lustig-Yaeger (University of Washington).

[6] NASA, *The LUVOIR Final Report*, 2019, NASA LUVOIR Mission Concept Study Team, https://asd.gsfc.nasa.gov/luvoir/reports/LUVOIR_FinalReport_2019-08-26.pdf.

[7] NASA, *Habitable Exoplanet Observatory Final Report*, 2019, NASA Habitable Exoplanet Observatory Study Team, https://www.jpl.nasa.gov/habex/pdf/HabEx-Final-Report-Public-Release.pdf.

[8] A. Misra, V. Mesdows, and D. Crisp, 2014, The effects of refraction on transit transmission spectroscopy: Application to Earth-like exoplanets, *Astrophysical Journal*, 792:61.

Both LUVOIR and HabEx teams have selected a suite of focal plane instruments that are aimed at both detailed characterization of exoplanets and more general observations. Table I.5 lists the capabilities of both LUVOIR and HabEx instruments. Central to discussing which of these missions is needed to achieve the science goals outlined by the science panels is the size of the telescope's primary mirror. The size of the primary mirror sets the angular resolution and the sensitivity achievable. The angular resolution scales with the diameter D, and the sensitivity scales as D^4 for background- and diffraction-limited imaging. For the goal of measuring exoearth biosignatures, a mirror sufficiently large to see the exoearth outside the coronagraph's inner working angle is required and needs to be large enough to measure the reflected light spectrum of the exoearth. An exoearth in the habitable zone of a G star at 10 parsecs has a magnitude of ~30 in the AB system, which alone suggests that a mirror with a collecting area of at least 6 meters is needed, and which is confirmed by the small number of candidates for small telescopes shown in Figure I.2. Table I.4 lists the estimated number of exoplanets detectable using any of the LUVOIR and HabEx configurations considered here (assuming $\eta_{Earth} = 0.24$).

Figure I.2 presents the expected number of exoearth candidates assuming a value of $\eta_{Earth} = 0.24$. During the span of the panel's work, it became clear that the value of η_{Earth} could be as much as a factor of 2.5× smaller, with a corresponding drop in the number of exoearth candidates (but see Bryson et al., 2020,[9] which supports a value of 0.24, while Gaudi et al., 2021, cite a range of 0.05 to 0.5).[10] The panel assumed that an extensive flagship mission would be capable of detecting an exoearth with essentially complete certainty. The number of detectable planets depends on $D^{1.97}$ and on $\eta_{Earth}^{0.96}$ using the formulation in Stark et al. (2019). If η_{Earth} proves to be as low as 0.05, then a mirror of inscribed size >=6 m is required to ensure detecting at least one exoearth. Figure I.3 shows how such a change in η_{Earth} will change the exoearth yield as a function of telescope size and coronagraph type. Figure I.3 also illustrates that coronagraph performance has not reached physical limits, and that recent progress in coronagraph performance is very encouraging.

TABLE I.4 The Numbers of Exoplanets of Various Types That Would Be Detectable Via Direct Imaging

Mission	Rocky + super-earth	HZ rocky + super-earth	Sub-neptune	Neptune	Jovian
Habex 4H	55	8	60	32	31
Habex 3.2S	23	4	40	56	
Habex 2.4H	19	3	27	30	
LUVOIR-A	230	54	210	92	117
LUVOIR-B	195	28	197	92	104

NOTE: These values assume $\eta_{Earth} = 0.24$ and were taken from the LUVOIR and HabEx study reports, which presented these same values. There is debate about whether η_{Earth} may be lower; the latest analysis performed by the Kepler team[11] is consistent with this estimate, but see also Gaudi et al. (2021), arXiv:2011.04703v.

[9] S. Bryson, M. Kunimoto, R.K. Kopparapu, J.L. Coughlin, W.J. Borucki, D.Koch, V.Silva Aguirre, et al., 2020, The occurrence of rocky habitable-zone planets around solar-like stars from Kepler data, *Astronomical Journal*, 161:36.

[10] Gaudi et al., 2021, arXiv:2011.04703v.

[11] S. Bryson, M. Kunimoto, R.K. Kopparapu, J.L. Coughlin, W.J. Borucki, D. Koch, V. Silva Aguirre, et al., 2020, The occurrence of rocky habitable-zone planets around solar-like stars from Kepler data, *Astronomical Journal*, 161:36.

TABLE I.5 Focal Plane Instrumentation for LUVOIR-A (Upper Table) and HabEx 4H (Below)

ECLIPS	
Coronagraph with imaging and imaging spectroscopy	
Bandpass	200–2000 nm
Contrast	1×10^{-10}
IWA	3.5 λ/D
OWA	64 λ/D
R ($\lambda/\Delta\lambda$)	Vis: 140 NIR: 70, 200

HDI	
Wide field imager with simultaneous UV/Vis and NIR coverage	
Bandpass	200–2500 nm
FoV	$3' \times 2'$
67 science filters + grism	
Nyquist sampled	
High-precision astrometry	

LUMOS	
UV/Vis multi-object spectrograph and FUV imager	
Bandpass	100–1000 nm
MOS FoV	$2' \times 2'$
Apertures	840 × 420
R ($\lambda/\Delta\lambda$)	500–50,000

POLLUX	
Point-source UV spectropolarimeter (European study for LUVOIR-A only)	
Bandpass	100–400 nm
R ($\lambda/\Delta\lambda$)	120,000
Circular + linear polarization	

	Coronagraph (HCG)	Starshade (SSI)	Workhorse Camera (HWC)	UV Spectograph (UVS)
Purpose	Exoplanet imaging and characterization	Exoplanet imaging and characterization	Multipurpose, wide-field imaging camera and spectograph for observatory science	High-resolution, UV imaging and spectroscopy for observatory science
Instrument Type	Vector Vortex charge 6 coronagraph with: - Raw contrast: 2.5×10^{-10} at the IWA - Δ mag limit = 26.5 - 20% instantaneous bandwidth - Imager and spectograph	52 m diameter starshade occulter with: - 76,600 km separation (Visible) - Raw contrast: 1×10^{-10} at the IWA - Δ mag limit = 26.5 - 107% instantaneous bandwidth - Imager and spectograph	Imager and spectograph	High-resolution imager and spectrograph
Channels	Visible: 0.45–0.975 μm - Imager + IFS with R = 140 Near-IR: 0.975–1.8 μm - Imager + IFS with R = 40	UV: 0.2–0.45 μm - Imager + grism with R = 7 Visible: 0.45–0.975 μm - Imager + IFS with R = 140 Near-IR: 0.975–1.8 μm - Imager + IFS with R = 40	Visible: 0.37–0.975 μm - Imager + grism with R = 1,000 Near-IR: 0.95–1.8 μm - Imager + grism with R = 1,000	UV: 115–320 nm (with 115–370 nm available at $R \leq$ 1,000) R = 60,000; 25,000; 12,000; 6,000; 3,000; 1,000; 500; imaging
Field of View	IWA: 2.4 λ/D = 62 mas at 0.5 μm OWA: 32 λ/D = 830 mas at 0.5 μm	IWA: 58 mas at 0.3–1.0 μm OWA: 6 arcsec (Vis. broadband imaging) OWA: 1 arcsec (Visible IFS)	3 x 3 arcmin^2	3 x 3 arcmin^2
Features	64 x 64 deformable mirrors (2) Low-order wavefront sensing and control	Formation flying, sensing, and control	Microshutter array for multi-object spectroscopy - 2 x 2 array, 171 x 365 apertures	Microshutter array for multi-object spectroscopy - 2 x 2 array, 171 x 365 apertures

NOTE: The baseline design for LUVOIR-B does not include POLLUX. HabEx 3.2S has no coronagraph and its HWC camera has a short wavelength cut-off of 0.32 micron. These tables are from the mission final reports.

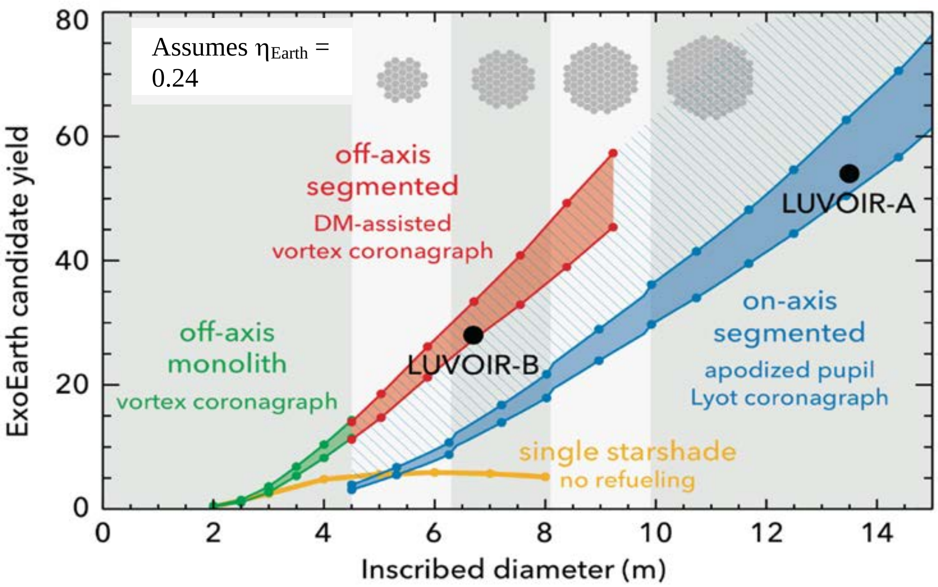


FIGURE I.2 The expected number of exoearth candidates assuming a value of η_{Earth} = 0.24.

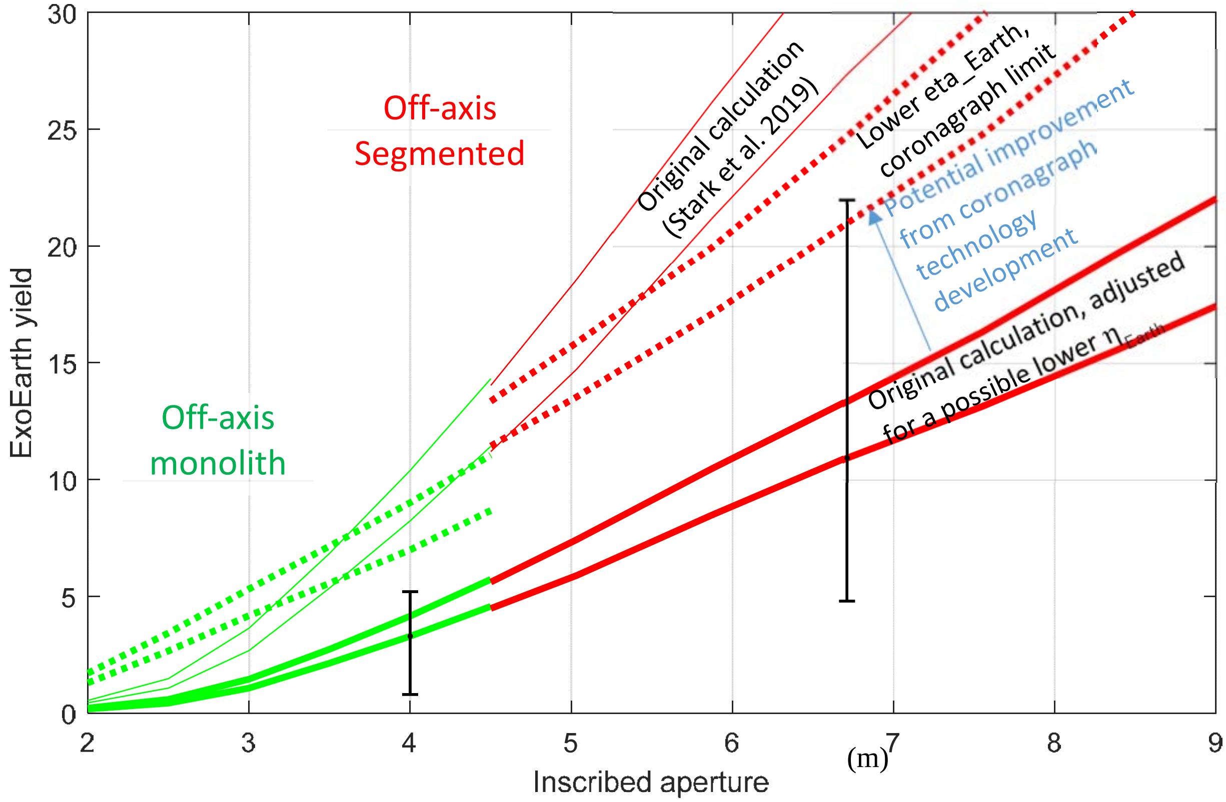


FIGURE I.2 Exoearth yields, showing the effects of a reduced η_{Earth} as well as potential improvement from future technology development in coronagraphs. The thin pair of curves labeled "Original calculation (Stark et al., 2019)" is the same as the red and green curves in Figure I.2. (The two curves span the range between "high-throughput" and "low-throughput" scenarios.) The thick curves at the bottom show the reduction based on the assumption that η_{Earth} is 2.5 times lower than what was assumed in the LUVOIR and HabEx reports. Black uncertainty bars, placed at the inscribed apertures for HabEx 4H and LUVOIR-B, represent uncertainties in η_{Earth}, exozodi, and astrophysical realizations as described in Stark et al. Dotted red lines represent theoretical limits of coronagraphs for the pessimistic 2.5 times lower η_{Earth} case. The blue arrow represents a gap between current coronagraph designs and theoretically achievable performance, which can be closed by continued coronagraph technology development, and make up for a smaller η_{Earth}. (New designs created after the LUVOIR and HabEx reports appear to already close about 40 percent of this gap; see Section I.3.5.) The panel used a similar methodology as Stark et al. (2019) to produce the modified curves shown here.

Other science cases also point to needing a telescope of at least 6 m. Figure I.4, taken from the LUVOIR team presentation to the panel, shows the number of quasi stellar objects (QSOs) detectable in the UV that could be used as background sources for tracing circumgalactic gas. Too few QSOs are accessible to HST for this technique to go beyond just showing the existence of such gas. Larger telescopes can detect more QSOs with adequate signal to noise to enable mapping of the circumgalactic medium, with Figure I.4 implying that D > 6 m will make a significant impact in this area. This impact comes not just in the form of more accessible QSOs but more importantly in the ability to study absorption lines in the spectra of QSOs at z ~ 0.5–1, which are essentially inaccessible to smaller telescopes. Using higher redshift QSOs enables study of the CGM around galaxies over nearly half the age of the universe. Many other exoplanet questions such as that posed by EASS-2: "What is the nature of individual planets, and which processes lead to their diversity?" also need D > 6 m (LUVOIR Report, p.

1-24).[12] Of the 12 "signature science cases" listed in **Error! Reference source not found.**, only 2 are cited as requiring a 15 m telescope. Four are cited as needing D >= 6.7 m, with the rest requiring 8 m. Some of the 8 m projects can be done using a slightly smaller telescope at the expense of needing more observing time. Based on these considerations and the expected yield of exoearths, the panel has set the minimum mirror size at 6 m. The panel notes that the time observing time difference between a telescope with a 6-meter collecting area and 6.7-meter collecting area is ~55 percent, which was judged an acceptable difference for these projects.

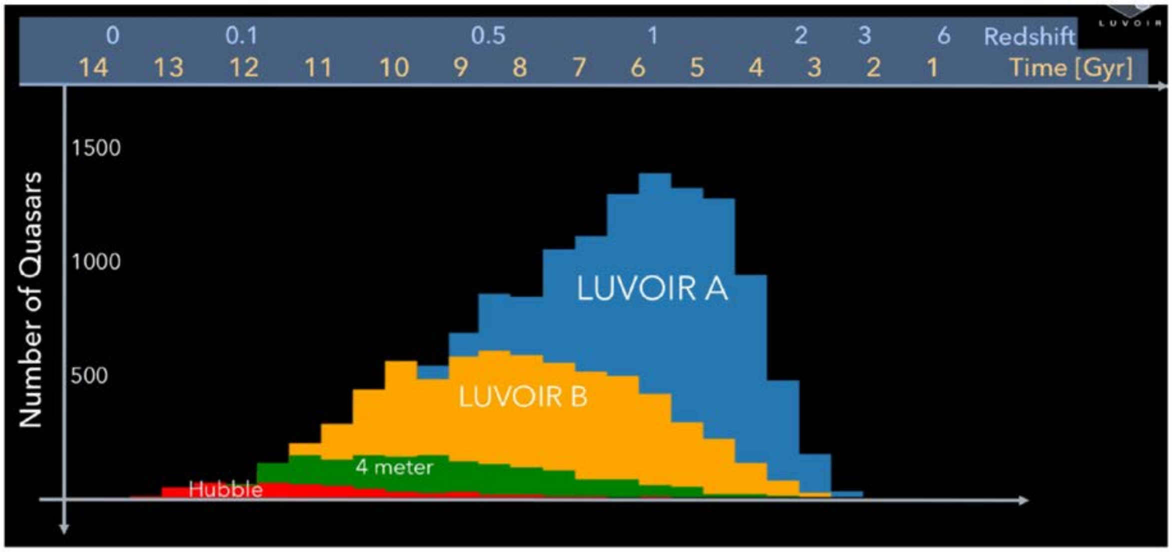


FIGURE I.4 Number of QSOs sightlines for use in measuring hot gas component of the universe. SOURCE: LUVOIR Team presentation to EOS-1 Panel. Courtesy of the LUVOIR Science and Technology Definition Team.

I.3.2 Flagship Costs

The two versions of LUVOIR, LUVOIR-A and LUVOIR-B, are estimated by their project teams to cost $16.0 billion and $12.2 billion in fiscal year (FY) 2020 dollars at a 70 percent confidence level. Because TRACE funding was limited, the panel asked the Aerospace Corporation to perform a TRACE analysis on only LUVOIR-B because LUVOIR-A seemed infeasible to complete during the next two decades without an unprecedented increase in the Astrophysics budget as judged from the team's report. Table I.6 summarizes the team and TRACE cost estimates for the three missions considered in detail by the panel. These cost estimates assume an optimal funding profile, and if development is funding-limited and stretched over a longer period than the 10–12 years assumed for these missions, the total costs will be significantly higher.

TABLE I.6 Comparison of Team and TRACE Cost Estimates ($FY 2020, 70% Confidence)

	Mission Team Cost Estimate	TRACE Cost Estimate
LUVOIR-B	$12.2 billion	$17 billion
HabEx 4H	$6.8 billion	$10.5 billion
HabEx 3.2S	$5.0 billion	$7.8 billion

[12] NASA, 2019, *The LUVOIR Final Report*, NASA LUVOIR Mission Concept Study Team, https://asd.gsfc.nasa.gov/luvoir/reports/LUVOIR_FinalReport_2019-08-26.pdf.

I.3.3 Technology Development Needs

All of these missions and the 6 m version as suggested by the panel will need significant technology development to reach the 10^{-10} contrast needed for direction detection of exoearths. These needs are highlighted by the Aerospace risk ratings of medium-high for LUVOIR-B, medium for HabEx 4H, and medium-low for HabEx 3.2S. For LUVOIR-B, Aerospace lists these areas for significant development:

- Large segmented mirror high-precision figure and stability;
- Coronagraph incorporates new technology for contrast improvements over WFIRST CGI;
- UV instrumentation requires fabrication improvements to optical coatings and detectors; and
- Large 48 m × 48 m sunshade requires consideration of stowage and deployment.

The equivalent list of HabEx 4H includes:

- Starshade to be scaled-up to 52 m diameter with integration of optical cover and solar array in deployable center disc;
- Coronagraph incorporates new technology for contrast improvements over WFIRST CGI;
- Telescope mirror with high-uniformity CTE and coating over 4 m monolithic, lightweighted primary mirror; and
- Life testing required for colloidal thruster to meet 10-year objective.

The list for HabEx 3.2S includes only the starshade and thruster risks from the HabEx 4H list. In addition to the items listed above that were noted by Aerospace, the panel found a number of other technology development items that are discussed in Sections I.3.4 and I.3.5.

The technology progress described earlier provides an excellent starting point for these missions. The comparison of HabEx 4H with HabEx 3.2S in the Aerospace risk assessment reveals that a starshade with its highly desirable properties of very small inner working angle and overall larger area of high contrast is not a high-risk item at the size scale needed for the HabEx telescopes but is rated as a medium risk with further development needed. A starshade for a larger telescopes would need to be larger and placed farther away from the telescope, which would increase the risk. Although the HabEx 4H observational plan includes a clever scheme to work around the limitation of needing refueling for a starshade to be positioned for many targets, the recent successes of refueling other satellites (Intelsat 901 was refueled by Northrop Grumman's MEV-1) could be applied to starshades, making them even more attractive. A space demonstration of a starshade, even at smaller scale than needed for an exoearth mission, would be valuable in retiring operational risks and as further proof beyond the small-scale ground tests of the efficacy of starshades.

Telescopes employing coronagraphic starlight suppression to achieve 10^{-10} contrast require a significant technology investment probably as large as $600 million ($FY 2020) by the start of phase A. Approximately $130 million of this funding would be needed for the high-contrast coronagraph instrument as estimated in the LUVOIR final report. Given the outstanding science discoveries that these exoplanet missions can accomplish but only with critical technologies that need to reach TRL 6 or above in the next 5 years, the panel suggests that NASA fund the technology tasks outlined in the LUVOIR and HabEx reports and do so over the next 5-year period, which would then flow into detailed mission architecture studies that could be completed before the next decadal survey, A key decision is the size of the primary mirror, and whether a starshade will be included in the mission. For a coronagraph-based mission, Figure 11-3 in the LUVOIR final report lays out a phased technology program that addresses the highest risk technology developments needed. This early funding of significant technology is not normal for a NASA flagship mission. However, given the significant cost uncertainties related to technology, a significantly more accurate overall flagship cost would result from this early technology development

roadmap, and would prevent getting well into mission design and development only to discover a significant issue that would incur a high cost to work around.

The technology roadmaps outlined by both LUVOIR and HabEx are critical to the ultimate determination of whether these exoplanet missions can accomplish their scientific goals. While the two mission concepts have some significant technology needs that are quite different (e.g., large sunshade for LUVIOR, starshade for HabEx), they do have a number of technology areas in common. The panel suggests that all 2020 decadal study high-risk technologies for these missions be reviewed for overlap such that a single technology roadmap can be described, costed, and scheduled with Figure 11-3 from the LUVOIR report illustrating at a top level what such a plan might look like. Such a plan might include key decisions points with oversight by a single program officer. A preliminary list of cross-program technologies that could be considered for this grand technology roadmap include ultra-stable structural composites, low-creep adhesives, CTE measuring techniques, milli-K thermal sensing, finite element model/test surface figure error/wavefront error model correlation. In addition, the panel suggests that development of starlight suppression technologies be continued, both starshades-based and coronagraph-based ones. There is significant overlap in the advancement of coronagraph-based technologies between LUVOIR and HabEx, and high-priority technologies to mature include better coronagraph architectures to increase science yield, adaptive wavefront control algorithms that are more efficient and improve tolerance to instabilities, vacuum-compatible deformable mirrors with more actuators, post-processing algorithms, and vacuum testing of the entire starlight suppression system. High-priority starshade technologies include demonstrating starshade petal accuracy and stability, performance modeling and validation, as well as demonstrating acceptably low scattered sunlight from petal edges. Ideally, this grand technology roadmap would be funded such that it could be fully accomplished within 5 years. Without studying each of the building blocks prior to the system-level testing, additional program risk remains.

To keep this technology roadmap focused, the panel suggests setting a strong, concise goal for a flagship: detect a suite of biosignatures on an exoearth, or show that they are rare in a statistically meaningful sense. The mission could start development before the end of the 2020s if the technologies and funding have met the panel's proposed criteria for technologies and budget (see section I.2, "Budgetary Considerations"). Because of the combination of the preceding challenges and opportunities, a commitment to the above goal rather than choosing a specific mission implementation is most appropriate now. A mid-decadal committee would review technology and scientific progress with NASA leading an effort to optimize the best mission point design in advance of that review.

I.3.4 Detailed Technology Development Comments—Telescopes Considerations

The panel is neither suggesting a preferred primary mirror configuration nor suggesting that monolithic primaries larger than 4 m are not feasible, but rather is indicating study needed to make an informed choice. For LUVOIR-style telescopes, three technology areas enable the science objectives: the high-contrast coronagraph, the ultra-stable segmented telescope that enables picometer-level wavefront and contrast stability, and the UV instrument technology. The development of these technologies to a level of TRL 6 prior to start of Phase A will enable the overall cost and schedule risk to be minimized. The LUVOIR mission study report has outlined the technology, engineering, and manufacturing needs and in general shows how the TRL in these areas could be improved to level needed to start mission development. Table 11-5 in the LUVOIR report summarizes the ultra-stable enabling technology development activities needed. Risks that do not seem to have been fully considered in the LUVOIR documentation related to these technology activities include the following (section numbers taken from the LUVOIR report):

- *11.2.2.1 System-Level Model Development and Validation*—The LUVOIR-B telescope has three times the number of mirror segments as JWST. The mechanical and thermal finite

element models created for JWST required significant amounts of computer power and very long run times. With the model ~3 times larger, a concern arises about the ability to solve mechanical and thermal problems (with < milli-K resolution) and perhaps hundreds of millions of degrees of freedom.

- *11.2.2.2 Thermal Sensing and Control Development*—The mission study report has shown that a very small solid mirror can be controlled in a small chamber to < 1 milli-K over long periods of time. The engineering study would benefit from including the most complicated areas where thermal sensing and control may not be straightforward; such as the internal and external perimeters of the primary mirror. This may be best addressed during the development of the segmented telescope system (see below) but also can be addressed with very detailed thermal and mechanical modeling.
- *11.2.2.3 Composite Material Process Development and Optimization*—Ultra stable telescopes need support structures that show very limited time-dependent delta L owing to temperature, moisture, and material creep effects. This latter term does not seem to be included in the study plan and may be as important as the other contributing factors. Other materials, such as invar, also have long-term material instability and are currently being assessed on Roman WFIRST for their impact on short-term stability. The LUVOIR design includes materials and components that are not as stable as the optical materials themselves; physically testing these materials and components may be needed.
- *11.2.2.4 Mirror Substrate*—The current plan does not include the creation of a 3D coefficient of thermal expansion (CTE) profile of each and every mirror (or in fact any single mirror). This may be necessary to get the predicted wavefront error (WFE) when very small temperature deltas are assumed. The substrate manufacturer currently does not thoroughly perform these tests.
- *11.2.2.12 Segmented Telescope System*—This will be the most critical TRL element because it will include many of the completed activities in Table 11-5 in the LUVOIR final report. Even though this activity is a subscale of the telescope system, the panel suggests that it be designed to include any/all areas where the greatest uncertainties may exist. All of the TRL items that will feed into this test need to be completed early enough and with as much schedule margin as possible to allow the Telescope System test to be thoroughly debugged and tested.

The HabEx mission study report identified two primary technology study items that will allow the telescope risk to be reduced significantly; demonstrating large mirror fabrication and obtaining the necessary mirror coating uniformity. The panel has identified other significant risk-related shortcomings that include lack of a full 3D map of CTE data, which would allow test and model comparisons to be made and understood; the effectiveness of their 1 g offloader system to fabricate their 0 g mirror figure; and the effects of long-term material stability not correctable by their laser metrology system (such as mirror figure time-dependent errors). There is no test plan to reduce the risk related to the effects of the lack of a 3D CTE map. The team contends that if the wavefront error requirement in a nominal thermal environment is met, then there is no significant risk. The problem is that the on-orbit environment(s) cannot easily or feasibly be duplicated on the ground (as determined with JWST) and therefore no model correlation or WFE validation will have been proven. This is a correctable flaw in the HabEx technical path via independent testing of CTE and performing model/test correlation. Another risk is related to the 1 g manufacturing and the need for actuators to achieve the correct mirror figure at 0 g. Figure 6.8-3 from the HabEx final report shows that errors would not be reduced sufficiently without actuators. Typical finite element model analyses are not likely to bring the errors below a factor of 2× above the stated HabEx requirement, which suggests that this is another area needing more attention. Mirror coating uniformity is adequately addressed in the HabEx report.

I.3.5 Detailed Technology Development Comments—Starlight Suppression Considerations

In high-contrast imaging, it is useful to distinguish between the planet:star flux ratio and the instrumental contrast. The flux ratio is a property of the astrophysical system—for example, Earth has approximately 10^{-10} brightness of the Sun. The instrumental contrast is (roughly) the ratio of the surface brightness of the residual halo of scattered light surrounding a star to the peak surface brightness of an unocculted star. Using post-processing to fit and remove residual starlight, it is often possible to detect and characterize a planet whose flux ratio is lower than the instrumental contrast, just as ground-based infrared instruments can detect objects much fainter than the bright sky. The exact post-processing benefit will depend on the stability of the instrument and telescope, driving toward ultra-stable observatory architectures. Both the LUVOIR and HabEx study reports included results from very good and detailed integrated modeling of the telescope and starlight suppression systems: dynamical Structural, Thermal, and Optical Performance (STOP) models, end-to-end diffraction and wavefront control loops, and so on. These models have a level of fidelity that goes well beyond prior studies of coronagraphic or starshade instruments (except for Roman WFIRST). The mission studies leverage the experience and some of the machinery of Roman WFIRST models, as well as Roman WFIRST hardware demonstrations. Both of these increase confidence in the coronagraphic and starshade starlight suppression technologies.

However, the TRL of some of the subsystems is still as low as 3. The reports describe coordinated plans to advance the technologies to TRL 6 prior to phase A, with a schedule and cost commensurate with past coronagraph technology development projects, including Roman WFIRST. Some of the key remaining risks are as follows:

- The gap between laboratory demonstrations of coronagraphs and LUVOIR/HabEx mission requirements is rapidly shrinking, but nonnegligible (see Table I.7). Contrast levels approaching 1×10^{-10} in broadband have been demonstrated, but with a simpler coronagraph than baselined for LUVOIR and HabEx, and for monolithic apertures. Currently, the deepest demonstrations for nonmonolithic apertures are order (1×10^{-9}) for Roman WFIRST coronagraphs (Hybrid Lyot and Shaped Pupil). For more aggressive coronagraphs, such as the Apodized Vortex that was chosen for LUVOIR-B, the demonstrations are of order (1×10^{-8}). Validated models predict that required coronagraph performance is achievable, which increases confidence that this gap will be closed soon. However, just as with any technology development project, there is always a small possibility of some unexpected limiting factor that escaped attention, because working at 1×10^{-10} contrast levels is still somewhat unexplored territory. (Possible examples: vector effects from segmentation edges, amplitude errors from reflectivity nonuniformity, sub-wavelength physics [if any], effects of dust or other contaminants, deformable mirror mechanical and electrical stability and reliability, etc.) Both LUVOIR and HabEx mission study reports present a good technology development plan to close this performance gap, and the panel emphasizes that a well-planned early technology development effort is absolutely critical for these missions.

TABLE I.7 Selected Laboratory Demonstrations of Coronagraphs

Aperture	Monochromatic Contrast	Broadband Contrast	Region	Coronagraph	Reference
Unobstructed	2×10^{-10}	4×10^{-10} (10% band)	3–8 λ/D, 360-degree FOV	Classical Lyot	Seo et al. (2019)[a]
Roman Space Telescope		1×10^{-9} (18% band)	3–9 λ/D, 360-degree FOV	HLC	Cady et al. (2017)[b]
LUVOIR-B	2×10^{-8}	4×10^{-8} (10% band)	6–10 λ/D, 60-degree FOV	Apodized Vortex	Llop-Sayson et al. (2020)[c]

[a] J. Llop-Sayson, G. Ruane, N. Jovanovic, D. Mawet, D. Echeverri, A.J. Edlorado Riggs, C.T. Coker, G. Morrissey, and H. Sun, 2019, The high-contrast spectroscopy testbed for segmented telescopes (HCST): New wavefront control demonstrations, *Proceedings of SPIE: Techniques and Instrumentation for Detection of Exoplanets IX* (S.B. Shaklan, ed.), vol. 11117, International Society of Optical Engineering (SPIE), Bellingham, WA.
[b] E. Cady, K. Balasubramanian, J. Gersh-Range, J. Kasdin, B. Kern, R. Lam, C. Mejia Prada, et al., 2017, Shaped pupil coronagraphy for WFIRST: High-contrast broadband testbed demonstration, *Proceedings of SPIE: Techniques and Instrumentation for Detection of Exoplanets VIII* (S. Shaklan, ed.), vol. 10400, International Society of Optical Engineering (SPIE), Bellingham, WA.
[c] J. Llop-Sayson, G. Ruane, D. Mawet, N. Jovanovic, C.T. Coker, J. Delorme, D. Echeverri, J. Fucik, A.J. Edorado Riggs, and J.K. Wallace, 2020, High-constrast demonstration of an apodized vortex coronagraph, *Astronomical Journal* 159(3):79–87.

- The amount of time and/or number of iterations required to achieve 10^{-10} contrast is based primarily on experience with WFIRST CGI modeling and demonstrations, which is at ~10^{-8} and 10^{-9} levels. It is possible that convergence to 10^{-10} will be slower than expected.

Fortunately, coronagraphic and wavefront control technologies are still improving, including several promising directions that mitigate the above risks:

- Advances in wavefront control algorithms (e.g., WP Pueyo, WP Kasdin) may relax the stability requirements of the telescope, potentially to JWST levels.
- Advances in post-processing algorithms may relax raw contrast requirements.
- Advances in coronagraph designs may relax telescope stability requirements, as well as enable reducing the size of the telescope and enable cost savings.
- The panel emphasizes that the technology is sufficiently mature to begin more detailed studies that would eventually lead to a LUVOIR- or HabEx-like mission now, and the above opportunities for improvement offer some combination of risk mitigation, increased science, and/or cost savings (e.g., by reducing telescope aperture without sacrificing exoplanet science). “Sufficiently mature” means that the interplay between mission architecture and starlight suppression techniques can be examined in greater depth effectively.

Observation: For either LUVOIR-B, a 6 m, HabEx-4H, or HabEx-3.2S to proceed to development, substantial investments in technology development will be required.

I.4 PROBES

At approximately $1 billion for development, a Probe mission line would fill the gap in cost between Explorers that are cost-capped at $150 million to $250 million and Flagships (typified by the >$3 billion Roman WFIRST and ~$10 billion JWST). This distribution of mission sizes leaves a gap of more than an order of magnitude in cost and scale between the large missions and the next category. A Probe line would provide an opportunity similar in scale to the smaller Great Observatories such as Spitzer and Compton. A Probe line could help ensure the availability of the pan-chromatic coverage needed because many problems in astrophysics benefit from multi-wavelength studies (e.g., WP Megeath) and a diversity of techniques such as ultra-high-resolution spectroscopy, polarimetry, and high time-resolution observations. WP Elvis also presents arguments supporting the need for probes. The panel heard presentations from all probes that included any UVOIR wavelengths as listed in Table I.3. The mapping of probes to science panel questions in Table I.3 shows that all of the probes that were considered by the panel can make a major or significant contribution to one or more of the science questions. Past missions that would have fallen into this cost category such as Spitzer and Kepler have made impressive contributions in a range of astrophysical areas. Probes also provide a mechanism for responding to new science opportunities on a more rapid time scale than is possible with larger missions.

The panel looked at the Planetary Science Division's New Frontiers program, similar in cost scope to what is envisioned for a Probe line, as a possible model. The panel envisions developing a probe line as soon as practical. Astrophysics will need to modify the program somewhat, but the EOS-1 Panel suggests that a Probe line would have the following characteristics. The line could support two Announcements of Opportunity (AOs) and two launches per decade. This cadence would be high enough to support PIs proposing but not being selected and coming back for a second try within a 10-year span. Examination of the probes presented to the panel suggests that a cost cap in the $1 billion to $1.5 billion range not including launch vehicle and science operations would be appropriate with the understanding that the total mission life cycle cost would be larger because of launch costs. Community participation in the science operations phase such as through a guest observer program is strongly suggested. NASA could consider incentivizing the use of new technologies needing demonstration in space by not counting the cost of such technologies against the total mission cost. The technologies to be allowed in this manner would be specified in the AO.

Given the quality of the probe missions presented to the panel, the panel found it tempting to specify the first Probe candidates. However, the panel considered the value of open competition for fostering new ideas and creativity, and suggests that the Probe line AOs not specify topical areas beyond the possible technology use mentioned above.

I.5 OVERALL PROGRAM BALANCE

The sections above lay out a strong case not only for exoplanet science but also for other astrophysics science goals and objectives. These were derived from the science panel inputs, and the panel has used these to help assess the various probe and flagship missions. In doing so, the panel considered program balance, which focuses on the need to ensure that opportunities are available for the whole astrophysics community to conduct science investigations, both on the ground and in space, and across all disciplines. With respect to the EOS-1 charge for electromagnetic observations from space, the current astrophysics program of record includes a plan for four Explorer missions and four missions of opportunity per decade and extremely large flagship missions that have been taking more than 20 years to complete. The flagships consume nearly half of the astrophysics budget for decades until they launch and start producing science results.

I.5.1 Explorer Program

Excellent science is best enabled by ensuring programmatic balance across small, medium, and flagship-class missions, with robust support for research, analysis, and theoretical underpinning of the science data returned by all of these missions. Flagship missions have the potential to enable unique scientific opportunities and can tackle “big questions” in science that are otherwise unachievable. However, the extremely slow cadence between flagships (>20 years in some cases) means that the time between new starts is a significant fraction of a career. This slow cadence means that a flagship’s scientific requirements are developed long before it can be deployed, and its technology is frozen into place well before launch, resulting in a potentially very long cadence between infusions of new technology. Thus, flagships can be slow to respond to new and emerging science questions and new technology. By ensuring programmatic balance across small, medium, and large missions, along with adequate support for scientific analysis of data from NASA missions, NASA can answer big, ambitious scientific challenges while still remaining nimble enough to take advantage of emerging scientific breakthroughs and new technology. The panel observes that for the reasons outlined below, the astronomical community is well served with the programmatic balance as currently practiced by NASA.

Smaller competed missions offer opportunities for a wide array of scientists to lead missions and have the potential to broaden and diversify the pool of scientific leaders (provided meaningful efforts are made to ensure full participation of the entire community). They provide more rapid results compared to flagships, helping to ensure that science is accomplished in a timely fashion and injecting frequent new experimental data to provide necessary feedback to theoretical work (which needs to be well-supported). This class of missions offers an opportunity for the science community to take a direct hand in setting priorities and leading the effort to answer key and emerging scientific questions.

Smaller missions such as Explorers and a new line of Probes provide an opportunity to demonstrate new technologies that can be used to break open new fields, provide multi-wavelength access, respond to new and emerging scientific discoveries, and reduce risk for more costly flagship missions. The instrumentation, observing strategies, and analysis techniques employed on smaller missions can serve as valuable test-beds and risk reduction measures for larger, costly flagships. For example, the Kepler mission, which originated as a Discovery mission, greatly expanded understanding of the number and diversity of planetary systems outside of our own solar system using the transit technique outside the confounding effects of Earth’s atmosphere. The Spitzer Space Telescope, equivalent in today’s dollars to a Probe-class mission, was creatively repurposed beyond its original scientific objectives to probe the thermal structure, chemistry, and atmospheric dynamics of extrasolar planets. Additionally, both the Spitzer Warm Mission and the NEOWISE Reactivation served to validate passive cooling techniques that will be employed by JWST.

Small missions also play a vital role in ensuring technological vitality. Many areas of astrophysics depend critically on the availability of specialized technologies that have limited commercial or military applications, meaning that they are kept viable only through the continued investment of NASA’s scientific programs. The more rapid cadence of smaller, competed missions compared to flagship missions allows these specialty technology areas to be maintained. Examples include far-infrared detectors, X-ray detectors and optics, polarimetry, and UV optics and detectors. Small missions are essential for keeping specialty astrophysical technologies alive between flagships that can be separated by decades.

Despite the important role they play in providing risk reduction and ensuring availability of key technologies, small missions currently face a different standard of risk than flagships. The current competed line of Explorer missions, along with Discovery and New Frontiers missions in the Planetary Science Division, are strictly cost-capped and limited to employing only technologies that are at or above Technology Readiness Level (TRL) 6 or above by the time of their Preliminary Design Reviews (PDRs). The Discovery and New Frontiers lines in the Planetary Science Division have recently somewhat ameliorated the concern that these lower cost missions are unreasonably restricted from using less mature technology by providing opportunities for ride-along technology demonstrations. For example, the Deep

Space Optical Communications (DSOC) package was offered with an incentive to proposing teams in the 2015 Discovery round. Offering similar incentivized technology demonstration opportunities for the Explorer and Probe lines in Astrophysics may offer some chance to mature technology and expand wavelength coverage and diversity of available techniques.

The emerging areas of SmallSats and CubeSats are gaining the attention of astronomers. SmallSats are being proposed to fill several key gaps in astrophysical research—namely, the monitoring of sources for weeks or months at time—and at wavelengths not accessible from the ground such as X-ray, ultraviolet, far-infrared, and low-frequency radio. Time-domain astronomy has been an enormous challenge for any flagship mission, because those telescopes are highly sought after and shared among hundreds of programs annually, making long-duration observing nearly impossible. Other science cases for SmallSats being developed now include a wide variety of astrophysical experiments, including exoplanets, stars, black holes and radio transients, galaxies, and multi-messenger astronomy. Achieving high-impact research with SmallSats is becoming increasingly feasible with advances in technologies such as precision pointing, compact sensitive detectors, and the miniaturization of propulsion systems.

I.5.2 Foundational Programs

The panel observes that several areas outside the usual mission development flow could have profound and far-reaching consequences if pursued. Servicing such as refueling could be beneficial, because one of the limitations in the use of starshades is the need for fuel to reposition the starshade. HabEx 4H devised a hybrid coronagraphy scheme to use the starshade only for the most important targets. Refueling the starshade might be preferable.

Assembly of structures in space might alleviate some of the issues with launching large telescopes, but current capabilities are far from what is needed for precision structures. Investment in developing this area could enable ambitious future missions that would be considered by future decadal surveys. More innovation may be needed to enable very large future telescopes in space. Funding some “blue sky” thought teams and including some engineering support is needed to reduce the cost of future large space telescopes. Management and funding are other areas that could benefit from innovative approaches.

J
Report of the Panel on Electromagnetic Observations from Space 2

EXECUTIVE SUMMARY

The Panel on Electromagnetic Observations from Space 2 (EOS2) was charged to review space mission opportunities involving electromagnetic observations across the full spectrum, excluding the narrow band from the near infrared (IR) to the near ultraviolet (UV). In addition, the panel was asked to consider potential new space programs in gravitational radiation and cosmic particles. The panel discussed a wide array of white papers submitted by the community that are germane to this charge.[1] In particular, the panel examined materials in support of two proposed Flagship mission concepts, Lynx and the Origins Space Telescope, that provide new capabilities for X-ray and far-IR observations, respectively, as well as a suite of Probe mission concepts covering a variety of fields. The panel requested and received Technical, Risk, and Cost Evaluations (TRACE) (see Appendix O) of both Lynx and Origins.

The panel resonated with three key considerations that emerged from the community's white papers:

- A panchromatic approach to the future of space astronomy is the only way to address many of the high-priority science questions of our times.[2]
- The current paradigm for selecting, funding, and managing Flagship missions will not support the simultaneous development and operation of the multiple observatories required to provide such panchromatic coverage. A radically different approach is needed.[3,4,5,6]
- As the community moves forward to address more detailed and specific questions, it would be beneficial for NASA to enable and plan for a new approach of coordinated programs, involving multiple missions on multiple platforms.[7,8]

---

[1] See Appendix A for the overall Astro2020 statement of task, for the set of panel descriptions that define the panels' tasks, and for additional instructions given to the panels by the steering committee.

[2] S.T. Megeath, L. Armus, M. Bentz, B. Binder, F. Civano, L. Corrales, D. Dragomir, et al., 2019, The legacy of the great observatories: Panchromatic coverage as a strategic goal for NASA astrophysics, white paper submitted to the Astro2020 Decadal Survey.

[3] J. Tumlinson, J. Arenberg, M. Mountain, L. Feinberg, J. Grunsfeld, K. Sembach, N. Levenson, J. O'Meara, and M. Postman, 2019, The next great observatories: How can we get there? white paper submitted to the Astro2020 Decadal Survey.

[4] J.A. Crooke, M. Bolcar, and J. Hylan, 2019, Funding strategy impacts and alternative funding approaches for NASA's future flagship mission developments, white paper submitted to the Astro2020 Decadal Survey.

[5] J. Hylan, M. Bolcar, and J. Crooke, Managing flagship missions to reduce cost and schedule, white paper submitted to the Astro2020 Decadal Survey.

[6] M. Smith, "Zurbuchen: JWST Will Not Launch in March 2021," Space Policy Online, last update June 10, 2020, Zurbuchen: JWST Will Not Launch in March 2021—SpacePolicyOnline.com.

[7] N.A. Levenson, L.J. Storrie-Lombardie, and B.J. Wilkes, 2019, Scientific advancement through flagship space missions, white paper submitted to the Astro2020 Decadal Survey.

[8] M. Elvis, J. Arenberg, D. Ballantyne, M. Bautz, C. Beichman, J. Booth, J. Buckley, et al., 2019, The case for probe-class NASA astrophysics missions, white paper submitted to the Astro2020 Decadal Survey.

The vision this panel offers in response to these issues involves three major components that are summarized below. Table J.3, which appears at the end of this report, indicates the connections between each of these components and the high-priority science questions that emerged from the Astro2020 Science Panels. The EOS2 vision addresses 23 of the 30 questions, at least in part, while 19 are addressed extensively and in detail, or are design drivers for the program.

1. *Joint X-Ray/Far IR Flagship Program:* While both Lynx, an advanced X-ray observatory, and Origins, an advanced far-IR observatory, are exciting concepts, neither an X-ray nor a far-IR mission alone is sufficient to address the most pressing science of Astro2020. A combined program involving both X-ray and far-IR components would be much more powerful. The panel envisions a coordinated Flagship program, built around the especially compelling theme of studying the "cosmic dance" between black holes and galaxies—the intricate relationship between the growth of black holes in the universe and the evolution of galaxies that form and evolve around them. Because many galaxies are obscured by dust, it takes the synergy of two distinct kinds of observations to peer into their central regions: A high-sensitivity and high-angular resolution X-ray imaging mission that can detect accretion onto the black holes themselves, and a far-IR spectroscopic mission that detects and pinpoints both the effects of the intense black hole radiation and the effects of star formation and evolution on galactic energetics. For the purposes of this report, these notional missions are called "Fire" and "Smoke," respectively. Fire and Smoke are based on the proposed flagship missions Lynx and Origins, respectively, but are scaled to fit into a single flagship program organized around the investigation of the cosmic dance science. While optimized for that central focus, however, they will still enable a broad array of other science highlighted by the science panels. This program is scientifically compelling and daring. While achievable, it poses significant technological challenges that will empower NASA to stretch U.S. capabilities well beyond the state of the art.
2. *Time Domain Astrophysics:* A coordinated program to ensure a continuous U.S. presence in space for the study of transient and time-variable phenomena in the universe. This could involve different platforms, ranging from a single probe to a suite of much smaller or medium size missions, potentially including foreign missions. The required capabilities include: All-sky monitoring at hard X-ray/gamma-ray energies, transient localization capability that can position events at the few arcsecond level, and fast slew and follow-up imaging and spectroscopy at ultraviolet, X-ray, and near-IR wavelengths. Much of this suite of capabilities exists now with the Neil Gehrels Swift Observatory, but it is essential that it be preserved and enhanced through new launch opportunities for the future, and that it be optimized to support the new science that will come from the Laser Interferometer Gravitational-Wave Observatory (LIGO), the Cherenkov Telescope Array (CTA), and IceCube.
3. *Early Universe Cosmology:* A mission designed to provide high-precision measurements of the polarization of the cosmic microwave background at a range of frequencies. This would complement major ground-based facilities in exploring several of the greatest mysteries of fundamental physics—inflation, dark matter, dark energy, and neutrino mass, as well as the growth of structure with cosmic time.

Last, the panel reemphasizes the importance of NASA maintaining a healthy balanced program of space astrophysics mission opportunities on all scales, from suborbital rocket and balloon experiments, through Astrophysics Pioneers and Explorers, and the new class of Probes. In addition, the panel endorses the currently envisioned NASA participation in the European Space Agency (ESA) upcoming major missions, the Advanced Telescope for High Energy Astrophysics (Athena) and the Laser Interferometer Space Antenna (LISA).

J.1 INTRODUCTION

Panchromatic (multi-wavelength and multi-messenger) observations are the only path to disentangle and decipher the nature and history of our complex universe. The realization of that goal has been the basis of twenty-first-century space astrophysics. No single telescope alone can answer all of the most pressing questions in the field, from the nature of the Big Bang to the emergence of life on planets.

The EOS2 panel is inherently the most panchromatic of the various program panels convened for the Astro2020 Decadal Survey. The charge to the panel was to review and evaluate a large suite of proposed space mission concepts designed to address astronomy and astrophysics questions primarily by means of radio, far-infrared, and high-energy electromagnetic observations from space. A wide variety of phenomena are uniquely observable in the bands under this purview, ranging from the dust and spectral line emission in galaxies prominent at millimeter and far-IR wavelengths (~15–500 μm), to the decays of radioactive nuclei visible at MeV gamma-ray energies. Although not reflected in its name, the panel was also charged with considering nonelectromagnetic investigations in space, such as those designed to detect relativistic particles and gravitational waves.

The panel reviewed a total of 55 white papers from the community covering a range of diverse topics. Proposed space-based missions included experiments devoted to GeV and MeV gamma rays, hard X rays, high-resolution X-ray imaging, spectroscopy, timing, and polarimetry, far IR, millimeter and MHz interferometry, the cosmic microwave background, cosmic rays, neutrinos and gravitational waves. Other white papers made the case for technology development, including cryocooler technology, advanced X-ray optics, heterodyne receivers, and fully active telescopes in space.

Several white papers addressed program balance between large and small missions and wavelength coverage. The case was made for the continued importance of Flagship missions, but with a need to control cost growth through early investment in development, and continuing assessment of technical and schedule realism. The wisdom of creating a new Probe class of missions was argued to close the gap between Flagships and Explorers. One white paper recounted the strong legacy of the Great Observatories, and it argued for panchromatic coverage as a strategic goal for NASA astrophysics.

The charge to the panel also included a review of the current status of the field. Over the past three decades, many space experiments in the EOS2 relevant wavebands have been developed and launched. A summary of missions currently operating, or approved for development, is provided in Table J.1. While the broad array of missions listed suggests that there is already a wealth of observational capability across the spectrum, those missions do not possess the appropriate mix of sensitivity, or the spatial and spectral resolution, necessary to address all of the observational needs for the next decade. As the white papers emphasized, the future of the field is ripe for further investment in new facilities with significantly enhanced capabilities.

This report is organized as follows: In Sections J.2 and J.3, the Flagship mission proposals that the panel reviewed, Lynx and Origins, are discussed with their TRACE analyses. In Section J.4, a rationale is provided for reformulating these mission concepts into a single Flagship program consisting of two notional missions, Fire and Smoke, which are jointly optimized for the study of the "cosmic dance," the complex interaction of galaxies and the giant black holes at their cores. A new approach is suggested for the co-development of these missions, within realistic NASA budget profiles, that would allow them to be operating contemporaneously. In Section J.5, the Probe class of missions is discussed, and arguments are presented for the development of two targeted Probe programs, one in Time Domain Astrophysics, and the other in Early Universe Cosmology. In Section J.6, the panel discusses the balance of mission sizes and considerations for on-orbit servicing. Section J.7 is focused on two ESA missions in development, with U.S. participation: Athena and LISA. Last, a summary and final thoughts are given in Section J.8.

TABLE J.1 EOS2-Related Missions Operating or in Development

Mission	Agency or Country	Capabilities	Spectral Coverage	Expected Launch

		Large		
Chandra	NASA	X-ray imaging and spectroscopy Transmission grating spectroscopy	0.2–10 keV 0.08–10 keV	
Fermi	NASA	γ-ray imaging and spectroscopy Spectroscopy	30 MeV–300 GeV 8 keV–30 MeV	
SOFIA	NASA/DLR	2.7 m telescope IR imaging and spectroscopy	0.3–1600 μm	
XMM-Newton	ESA	X-ray imaging and spectroscopy Reflection grating spectroscopy UV/visible monitor	0.15–12 keV 0.33–2.5 keV 170–650 nm	
INTEGRAL	ESA	X- and γ-ray imaging and spectroscopy Visible monitor	3–35 keV; 15 keV–10 MeV 500–850 nm	
SRG	DLR/Russia	X-ray imaging and spectroscopy	0.2–10 keV; 5–30 keV	
HXMT	China	X-ray imaging and spectroscopy γ-ray monitoring	20–250 keV; 5–30 keV; 1–15 keV 0.2–23 MeV	
DAMPE	China	γ- and cosmic ray imaging and spectroscopy	5 GeV–10 TeV; 100 GeV–100 TeV	
		Medium		
SWIFT	NASA	X- and γ- ray imaging and spectroscopy UV/visible imaging	0.2–10 keV; 15–150 keV 170–650 nm	
ASTROSAT	India	X-ray imaging and spectroscopy UV/visible Visible	0.3–100 keV 200–300 nm; 130–180 nm 320–550 nm	
ISS-MAXI	JAXA	X-ray imaging and spectroscopy	2–30 keV; 0.5–12 keV	
GECAM	China	X- and γ- ray all sky monitor	6 keV–5 MeV	
		Small		
NuSTAR	NASA	X-ray imaging and spectroscopy	3–79 keV	
		Mission of Opportunity		
ISS-NICER	NASA	X-ray timing and spectroscopy	0.2–12 keV	
		Approved		
JWST	NASA	IR imaging and spectroscopy	0.6–28.3 μm	2021
GUSTO	NASA	IR high-resolution spectroscopy	63, 158, and 205 μm	2021
IXPE	NASA	X-ray polarimetry	2–8 keV	2021
SVOM	China/France	X- and γ- ray imaging and spectroscopy Visible imaging	0.3–10 keV; 4–150 keV 15 keV–5 MeV 400–950 nm	2022
EP	China/DLR	X-ray imaging and spectroscopy	0.5–5 keV; 0.3–10 keV	2022
XRISM	Japan	X-ray imaging and spectroscopy	0.4–13 keV; 0.3–12 keV	2022
SPHEREx	NASA	IR spectroscopy	0.75–5 μm	2024
HERD	China/ESA member states	γ- rays and electrons Cosmic rays	Tens of GeV–10 TeV Up to PeV	2025 (?)
eXTP	China/ESA member states	X-ray imaging Polarimetry Spectroscopy	2–50 keV 2–10 keV 0.5–10 keV; 6–10 keV	2027
ARIEL/CASE	ESA/NASA	IR spectroscopy Visible/IR photometry	1.25–7.8 μm 0.5–0.55 μm; 0.8–1.0 μm; 1.0–1.2 μm	2029
Athena	ESA	X-ray imaging and spectroscopy	0.3–10 keV; 0.1–12 keV	Early 2030s
LISA	ESA	Gravitational waves	2×10^{-5}–3×10^{-2} Hz	2034

J.2 PROPOSED X-RAY FLAGSHIP MISSION CONCEPT: LYNX

Lynx is one of the four Flagship concepts studied by NASA in preparation for the Astro2020 Decadal Survey.[9] The mission concept is based on the X-Ray Surveyor notional mission envisioned in the NASA Astrophysics Roadmap *Enduring Quests, Daring Visions*.[10] The Lynx observatory would operate in the 0.2–10 keV energy band with 100 times higher sensitivity than the Chandra X-Ray Observatory over a much larger field of view. The mission technical requirements are defined by the three scientific pillars described below, that map directly onto many of the Key Science Questions and Discovery Areas from the Astro2020 Science Panels, especially the panels on Compact Objects and Energetic Phenomena, Cosmology, Galaxies, Interstellar Medium and Star and Planet Formation, and Stars, the Sun, and Stellar Populations.

Pillar 1—The Dawn of Black Holes. Discover massive black holes (BHs) formed in the very first galaxies ($z \sim 10$) and determine the mechanism(s) by which they were able to so quickly assemble into the super massive black holes (SMBHs) seen at lower redshift ($z \sim 6$). While the host galaxies of the seed BHs will be found and characterized in deep optical and IR surveys obtained either with JWST or the subsequent Roman Observatory, only Lynx can reach the X-ray flux limits required to detect BHs (Figure J.1).

Pillar 2—The Invisible Drivers of Galaxy Formation and Evolution. Characterize the diffuse baryon population of galactic halos, observe the effects of AGN feedback on galaxies, determine the state of the gas feeding the central BH, and measure the energetics and mechanics of the resulting outflows.

Pillar 3—The Energetic Side of Stellar Evolution and Stellar Ecosystems. Study X-ray emission associated with stellar birth, evolution, and death over the entire initial mass range. Characterize stellar coronae to address crucial questions on planet habitability, owing to potential lethal effects of coronal activity on life on planets orbiting close to their stars. Study compact stars through surveys of X-ray binaries and supernova remnants in the Milky Way and other nearby galaxies.

Lynx would also play a major role in time domain/multi-messenger astrophysics by following up LIGO merger events, studying X-ray chirps in merging SMBH systems, and monitoring the evolution of tidal disruption events.

J.2.1 Instrumentation Required for Lynx

The Lynx Mirror Assembly (LMA) technical requirements flow from Pillar 1. The detection of BH seeds at $z \sim 10$ translates into a detection flux limit of 1×10^{-19} erg cm^{-2} s^{-1} for a 10^4 $M_{\odot}$ BH accreting at the Eddington limit, which is ~100 times fainter than the Chandra deep fields. To meet this requirement, the LMA must achieve sub-arcsecond (~0.5" or better) angular resolution in the energy range 0.2–10 keV, with 50 times the Chandra effective area, over a factor of 10 larger field of view (FOV) (22' diameter).

[9] NASA Marshall Space Flight Center, *Lynx X-Ray Observatory Concept Study Report*, 2019, NASA Science and Technology Definition Team, Huntsville, AL, https://wwwastro.msfc.nasa.gov/lynx/docs/LynxConceptStudy.pdf.

[10] NASA, 2013, *Enduring Quests, Daring Visions: NASA Astrophysics in the Next Three Decades*, NASA Astrophysics Subcommittee, Washington, DC, https://science.nasa.gov/science-red/s3fs-public/atoms/files/secure-Astrophysics_Roadmap_2013_0.pdf.

Permission Pending

FIGURE J.1 Representation of SMBH formation models, showing the sensitivity required to detect BHs early in their evolution (z > 10). Athena (see description in Section J.7.1) begins to detect such sources only when the BHs reach sizes of 10^7 $M_\odot$. Lynx can see them much earlier in their history. SOURCE: Lynx concept study.

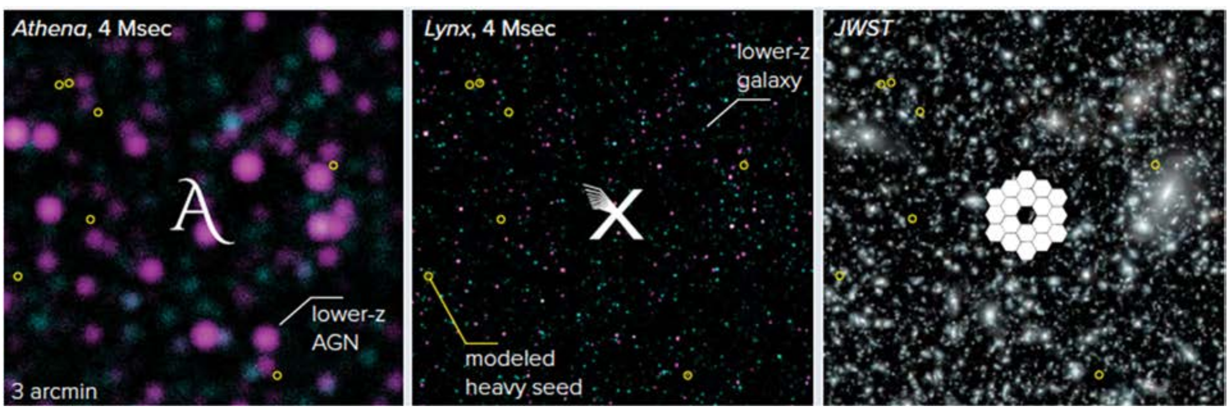


FIGURE J.2 Simulated deep surveys of 3' × 3' (2 percent of the total area) fields. *Left:* Athena (5" resolution). *Middle:* Lynx (0.5"). *Right:* JWST (0.1"). Lynx would not be affected by source confusion. Every X-ray source in the Lynx FOV can be uniquely identified with its host galaxy. SOURCE: Lynx concept study. Courtesy of the Lynx Team.

The High-Definition X-Ray Imager (HDXI), the Lynx wide-field imager, would be a silicon detector array with ~100 eV spectral resolution and 0.3" pixels (0.2–10 keV); it would provide good sensitivity and spatial resolution across the FOV and would exploit the full imaging potential of the optics (Figure J.2).

The Lynx *X-Ray Microcalorimeter (LXM)* would provide high-resolution imaging spectroscopy of point and extended sources using large arrays of microcalorimeters. It would comprise three different arrays: (1) Main Array: FOV 5', pixel size 1", energy resolution ~3 eV (*R* ~ 2000 at 6 keV); (2) Enhanced Main Array: FOV 1', pixel size 0.5", energy resolution ~2 eV (*R* ~ 3000 at 6 keV); (3) Ultra High Energy Resolution Array: FOV 1', pixel size 1", energy resolution ~0.3 eV (*R* ~ 2000 at 0.6 keV).

The Lynx X-Ray Grating Spectrometer (XGS) would provide an order of magnitude higher spectral resolving power (*R* ~ 7500 over the whole energy range) than the Chandra and XMM-Newton grating instruments, and greater than 500 times higher throughput at soft X-ray energies.

In summary, Lynx would provide several orders of magnitude in sensitivity enhancements for X-ray imaging and spectroscopy over both Chandra and ESA's (approved) Athena X-ray mission. While Athena matches the collecting area of Lynx, the 5" goal for Athena optics limits point source detection sensitivity (Figure J.2), which precludes the study of BH formation and evolution in the high-redshift universe. Athena's combination of moderate angular and (microcalorimeter-like) energy resolution is sufficient for point and significantly extended sources, but the Lynx combination of high throughput, 10 times better angular resolution, and ultra-high spectral resolution, is crucial for detailed studies of galactic and circumgalactic environments.

J.2.2 Technology Drivers and Associated Risks

The primary Lynx technology driver is the development of the X-ray optics. These require angular resolution ~0.5" (HPD) and large effective area (2 m^2 at 1 keV). The Lynx team selected the most mature technology, Silicon Meta-Shell Optics, for the Design Reference Mission (DRM), while maintaining full shell and adjustable segmented optics as risk reducing, potential "breakthrough" alternatives. The baseline approach exploits monocrystalline silicon segments and a highly modular design that achieves the requisite area and angular resolution by integrating tens of thousands of mirror segments into hundreds of mirror modules, all requiring alignments to a fraction of an arcsecond. The Silicon Meta-Shell approach has been demonstrated with mirror pairs mounted to 1.3" (HPD) at 4.5 keV, which, when corrected for gravity sag, suggests that sub-arcsecond performance is indeed achievable.

Currently, this technology is at Technology Readiness Level (TRL) 3, with good documented progress on the fabrication, coating, alignment, and bonding tasks required to reach TRL 4. Nevertheless, the highly modular approaches required to fabricate a complete optic, while maintaining both very high angular resolution and very large area, pose a major challenge and add considerable risk to the program. The mirror segments and modules need to be manufactured on an unprecedentedly large scale, while utilizing rapid, reproducible precision fabrication and optical alignment processes. The Lynx team clearly recognizes this industrialization challenge and the need for tackling it early, as evidenced by publications and by a recently established contract with industry partners experienced in large-scale fabrication.

J.2.3 TRACE Analysis of the Lynx Proposal

A TRACE analysis of the Lynx program plan was performed by the Aerospace Corporation to provide an independent assessment of the cost, schedule, and risk baseline. The TRACE assessment indicated that the Lynx project team's cost estimate of $6.2 billion ($ fiscal year [FY] 2020) has ~11 percent probability of not being exceeded. The TRACE cost estimate at 70 percent probability is $9.0 billion ($FY 2020), 45 percent higher. Additionally, the TRACE analysis found the Lynx schedule estimates for both Pre-Phase A to Phase A and for Phase B through launch to have a probability of not being exceeded of 22–25 percent. A 70 percent confidence schedule was estimated to require 6.75 years of technology development and Phase A formulation, and 11.5 years for Phase B–D development and launch, compared to the *Lynx* team estimates of 5 years for technology development and Phase A formulation, and 10 years for Phase B–D development and launch.

The primary differences between the TRACE analysis and that of the Lynx team were associated with the technology development and the flight build and integration of the LMA, which was assessed to be the highest risk area in the TRACE report. This assessment of risk is consistent with the opinions of this panel. As discussed above, the Lynx team has proposed continuing development of three independent X-ray mirror technologies. The panel agrees that the Lynx team has underestimated the complexity of

industrialization and integration/alignment of the optics, but also believes some savings relative to the TRACE estimate could be achieved by downselecting among the available technologies earlier. The other driver for the difference between the TRACE and Lynx estimates is in the area of reserves and margin, both during the Pre-Phase A and Phase A technology development and formulation phases, and during Phases B–D. When excluding reserves from both estimates, the Lynx subsystem development estimates are consistent with 40 percent confidence in the TRACE analysis.

Overall, the $2.8 billion increase in the total ($FY 2020) cost in the TRACE estimate contained an additional ~$800 million for technology development, including reserves, and $1.5 billion in Phase B–E reserves for potential schedule underscope and complexity growth, plus additional increases for management, systems engineering, and mission assurance. Based on the actuals used to perform the TRACE analysis, this panel believes the higher estimates for the Lynx cost and schedule are likely to be closer to reality than the Lynx team estimates. Additionally, implementing either the Lynx team $6.2 billion ($FY 2020), 15-year program or the TRACE $9 billion ($FY 2020), 19-year program, would require peak-year funding in excess of $1 billion ($ real year [RY]) for multiple years. Assuming the $500 million ($750 million) per year nominal (aspirational) NASA Astrophysics budget for new missions that was presented to the panel, an additional 3+ years would be required.

J.3 PROPOSED FAR INFRARED FLAGSHIP MISSION CONCEPT: ORIGINS SPACE TELESCOPE

The Origins Space Telescope is the second major Flagship mission concept that falls within the purview of the EOS2 panel.[11] The mission concept is based on the Far-IR Surveyor notional mission envisioned in the NASA Astrophysics Roadmap *Enduring Quests, Daring Visions*.[12] It would incorporate a 5.9 m cryogenic space telescope, operating from 2.8 to 588 μm, and deliver > 1000 times higher sensitivity than previous far IR/submillimeter missions. Origins science would address many of the key science questions and discovery areas identified by the Astro2020 Science Panels, especially the panels on Galaxies; the Interstellar Medium and Star and Planet Formation; Stars, the Sun, and Stellar Populations; and Exoplanets, Astrobiology, and the Solar System. Origins has three main science themes, described below, of broad interest to scientists and laypersons alike. The Origins mission core science cannot be addressed with complementary observations in other wavelength bands.

Theme 1—How do galaxies form stars, make metals, and grow their central SMBHs from reionization to today? The cooling processes central to the earliest phases of galaxy formation are largely facilitated by line emission observable in the far IR. Origins would enable spectroscopic observations in this band that provide an especially powerful probe of both BH irradiation and star formation. Both processes are well traced by the rich collection of far-IR line emission from various atoms, ions, and molecules (see Figure J.3). Origins traces the assembly and growth of galaxies, stellar populations, and gas phase metallicities through far-infrared spectroscopy. It also measures the gas inflowing to and outflowing from the galaxy, driven primarily by supernovae and AGN activity. These velocity-resolved spectra of species, such as OH, thereby trace the link between the buildup of BH mass and stars in galaxies across cosmic time.

Theme 2—How do the conditions for habitability develop during the process of planet formation? Stars form within molecular clouds through accretion-disk-like structures onto protostellar cores. Planets form from the residual protoplanetary disk. However, key questions remain about the role of hydrogen in gas giant formation, the role of water in the formation of habitable planets, and how the major building

[11] NASA Goddard Space Flight Center, 2019, *Origins Space Telescope Mission Concept Study Report*, Astrophysics Science Division, Greenbelt, MD, https://asd.gsfc.nasa.gov/firs/docs/OriginsVolume1MissionConceptStudyReport25Aug2020.pdf.

[12] NASA, *Enduring Quests, Daring Visions: NASA Astrophysics in the Next Three Decades*, 2013, NASA Astrophysics Subcommittee, Washington, DC, https://science.nasa.gov/science-red/s3fs-public/atoms/files/secure-Astrophysics_Roadmap_2013_0.pdf.

blocks of life get delivered to habitable planets. Origins would address these questions again through far IR spectroscopy. Line strengths trace gas mass and excitation, and velocity resolved observations tomographically place the line emission radially in the protoplanetary disk assuming Keplerian orbits (see Figure J.3).

Theme 3—Do planets orbiting M-dwarf stars support life? The search for life on exoplanets involves spectroscopy of the atmospheres in the “habitable zone” around main sequence stars. Measuring the abundances of species that can be linked with biology (e.g., O_2, O_3, CH_4, CO_2, H_2O, N_2O) that are clearly out of chemical equilibrium will be a signal for life. Origins would contribute to this field through transit spectroscopy in the mid-IR for Earth-like planets orbiting M or K dwarf stars. Transits by lower mass main sequence stars have greater depth, both owing to the relatively low flux from the parent star, and the relatively small radius of the habitable zone for M/K dwarf stars.

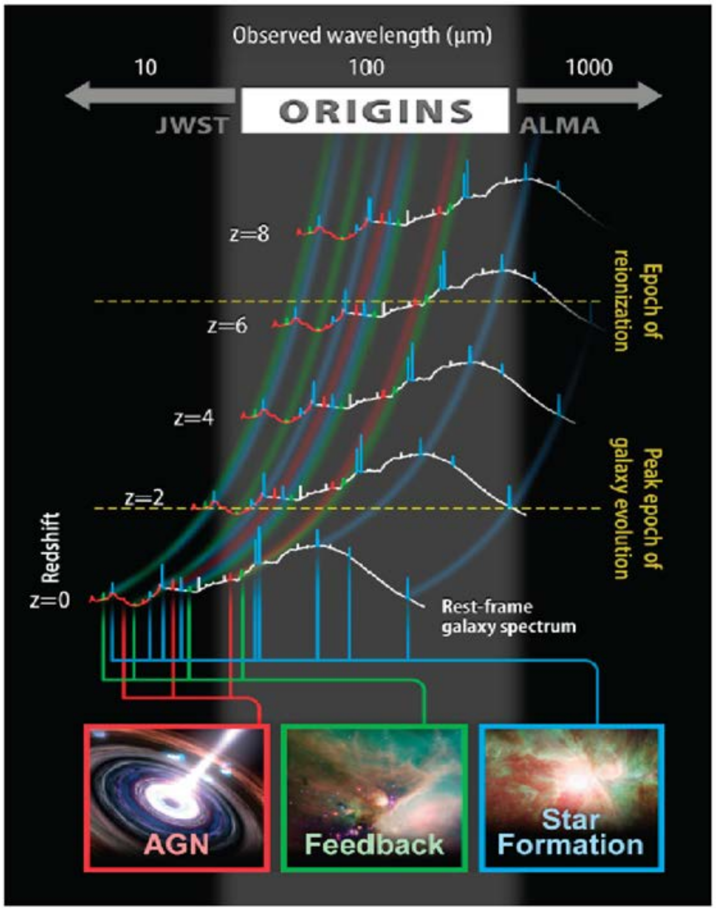
J.3.1 Instrumentation Required for Origins

The Origins Survey Spectrometer (OSS) would be a very broadband (25–588 μm) long slit grating spectrometer with resolving power R = 300, sufficient to separate line from continuum in the far-IR fine structure lines. Its sensitivity would allow the measurement of primary diagnostic lines for star formation from galaxies in the epoch of reionization at z > 6, as well as lines such as [OIV] 25.9 μm and [NeV] 14.3, 24.3 μm that measure both BH mass and accretion rates in the important 8–10 redshift interval when 10^4–10^5 $M_\odot$ primordial seed BHs grew by accretion to the $>10^6$ $M_\odot$ BH cores of the first galaxies. OSS would also include a Fourier transform spectrometer (FTS) placed in front of the grating to achieve resolving powers up to 43,000*(112 μm/λ), enabling both velocity-resolved spectra of galactic inflows and outflows (require R > 5000), and velocity resolved tomography of spectral lines from proto-planetary disks. A very-high-resolution “Etalon” stage (R ~325,000*[112 μm/λ]) would resolve protoplanetary disk spectral lines, enabling detailed studies of disk properties and the deposition of critical materials for the build-up of terrestrial planets.

The Far-IR Imager Polarimeter (FIP) would be an imaging polarimeter operating at two bands, 50 and 250 μm. With 8000 pixels, it would be capable of wide field (> 1000 deg^2), diffraction-imited imaging that would address a variety of important astrophysical topics from the evolution of star formation over cosmic time to transient follow-up and monitoring.

The Mid-infrared Spectrometer Camera Transit Spectrometer (MISC-T) would be a low-resolution (R = 50 to 300) imaging spectrometer delivering simultaneous spectra from 2.8 to 20 μm. Its state-of-the-art detectors and pupil densification would achieve the 5 ppm precision and stability necessary to detect the spectral lines that are biosignatures from exoplanets transiting main-sequence M and K stars.

In summary, Origins would provide unique or substantially enhanced capability with respect to all previous and planned FIR and submillimeter wave facilities. The OSS would be orders of magnitude more sensitive for spectral line surveys of high-redshift galaxies—1000 times more sensitive than Herschel. The only other proposed mission in this wavelength band was the Space Infrared Telescope for Cosmology and Astrophysics (SPICA), a 2.5 m, cold telescope that was a candidate for ESA's M5 selection, but was canceled before selection. The Atacama Large Millimeter Array (ALMA) can directly resolve protoplanetary disks, but Earth’s atmosphere prevents ALMA observations of both HD and neutral oxygen, and limits water observations to three high excitation lines. In the area of transit spectroscopy of low-mass stars, the MISC-T would have significantly higher stability than JWST, in the crucial 5–10 μm band.



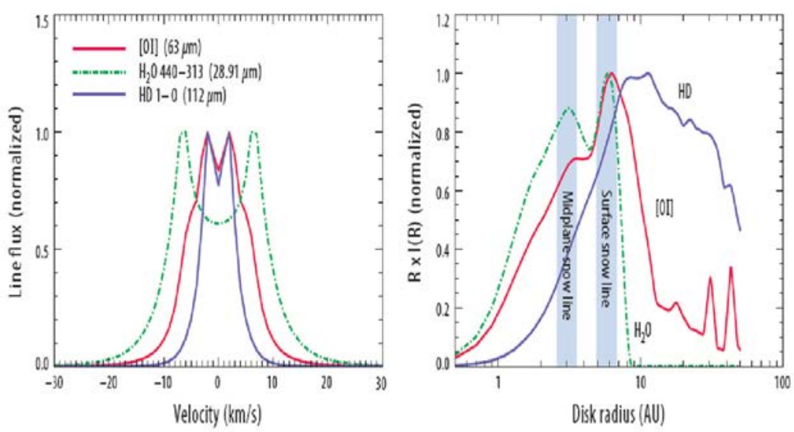


FIGURE J.3 *Top*: Infrared Space Observatory (ISO) far-IR spectrum of the Circinus galaxy showing spectral features detectable by Origins OSS at redshifts up to z = 8. The color coding indicates lines that are sensitive to AGN activity, feedback, and star formation. *Bottom*: Line-tomography example from protoplanetary disks. Assuming Keplerian orbits, the velocity resolved spectra of protoplanetary disks (left) reveal the location of important gas phase building blocks for planets through their line profiles (right). SOURCE: NASA Goddard Space Flight Center, 2019, *Origins Space Telescope Mission Concept Study Report*, Astrophysics Science Division, Greenbelt, MD. Courtesy of M. Meixner et al., 2019, arXiv:1912.06213. Reproduced with permission.

J.3.2 Technology Drivers and Associated Risks

The Origins concept would incorporate a 5.9 m primary mirror, cooled to 4.5 K, the requisite low temperature for reducing the FIR thermal photon background that would otherwise limit system sensitivity. The FIR focal plane employs superconducting transition edge sensor (TES) bolometers with multiplexed SQUID readout. The mid and near IR employs Si:As and HgCdTe arrays with an integrated readout. The Origins team identified the primary mirror, the cryogenic system, and the detectors as the primary technical challenges in the mission, and the panel agrees with that assessment. The TES bolometer arrays are of principal concern, given the large number of pixels required for the OSS, and the high degree of multiplexing involved. While prototype arrays with close to the required characteristics have been fabricated, these are not available at the production rate and high yield required to limit cost risk.

J.3.3 TRACE Analysis of the Origins Proposal

A TRACE analysis of the Origins program plan was also performed by Aerospace Corp to provide an independent assessment of the cost, schedule, and risk baseline. The TRACE assessment indicated that the Origins project team's cost estimate of $7.4 billion ($FY 2020) has ~25 percent probability of not being exceeded. The TRACE cost estimate at 70 percent probability is $10.6 billion ($FY 2020), 43 percent higher. Additionally, the TRACE analysis found that the Origins schedule estimates for both Pre-Phase A to Phase A and for Phase B through launch have a probability of not being exceeded of 42–47 percent. A 70 percent confidence schedule was estimated to require 6.25 years of technology development and Phase A formulation, and 9.3 years for Phase B–D development and launch, compared to the Origins estimates of 5.25 years for technology development and Phase A formulation, and 8.3 years for Phase B–D development and launch.

The primary differences between the TRACE analysis and that of the Origins team were in threat and reserve estimates, driving $56 million ($FY 2020) of the $88 million ($FY 2020) increase of the TRACE Pre-Phase A/Phase A estimate, and $2.75 billion ($FY 2020) of the $3.1 billion ($FY 2020) increase of the TRACE Phase B–E estimate. The schedule discrepancy between the two analyses was driven by the OSS instrument development and integration time, that the TRACE team estimated would require an additional ~1.5 years. OSS development was also identified as the critical path by the Origins team.

Although the allocation of reserves drove the TRACE analysis to predict a higher cost for Origins, the panel notes that the TRACE analysis predicted lower costs for the OSS and FIP instruments as compared to the Origins team estimates. The large difference in threat and reserve estimates is likely driven, at least in part, by the ~2000 kg difference in TRACE (higher) versus Origins team maximum expected value (MEV) mass. Overall, this panel expects that with careful management the cost to implement Origins is likely somewhere between the Origins team estimate and the TRACE estimate. As in the case of Lynx, implementing Origins within the schedules defined by either the project team or the TRACE analysis will require far in excess of the $500 million ($750 million) per year nominal (aspirational) NASA Astrophysics budget for new missions. The TRACE technically limited funding profile for Origins would require peak-year funding in excess of $1.45 billion ($RY) for multiple years. Executing this program would require reformulating the program plan to fit within a reasonable maximum annual funding limit. This would likely cause a significant increase to the duration of Phases B–D.

J.4 THE COSMIC DANCE VISION: A JOINT X-RAY/FAR-IR FLAGSHIP PROGRAM

As the panel reviewed the science of Lynx and Origins, it became clear that the missions complemented one another strongly. Neither mission by itself can address the full range of key science questions identified by the Astro2020 Science Panels, but together, they provide the required capabilities. When X-ray and FIR observations are both available, the whole is definitely more than the sum of its parts.

However, as the TRACE analysis clearly demonstrated, if NASA were to pursue these two missions in series, the launches would be > 20 years apart, seriously delaying—if not inhibiting—the full science return. *In the unanimous judgment of this panel, it is more important and more scientifically valuable to have contemporaneous advanced capability in both the X-ray and the FIR bands, than it is to have the full capability of either Lynx or Origins by itself.* Inevitably, some reoptimization would be necessary. Such an effort would focus on a *central scientific theme*, to enable the difficult decisions as to which elements of the present observatory designs are essential to maintain. Clearly, fitting two Flagships into one program cannot be accomplished without narrowing of scope to maintain cost reality.

The synergy between powerful X-ray and FIR observations is particularly strong in the study of the fundamental and intricate cosmic dance between star formation in galaxies, and the growth of their central BHs from the earliest times (see Section D-Q3). Each regulates the other, and their mutual evolution is an ecosystem that must be studied in its entirety from formation epoch to the present. This interplay is indeed one of the three pillars of Lynx, and one of the three themes of Origins. The X-rays detect the "fire," the central source of energy produced by accretion onto the BH, while the FIR detects the "smoke," the effect on the surrounding environment owing to the central irradiation, and its consequences on the galactic star formation and evolution. Because of the importance of this science, and the fact that it so strongly benefits from contemporaneous X-ray and FIR observations, the panel has chosen this to be the central defining theme of a joint X-ray/FIR program.

The cosmic dance began in the first billion years of the universe with first light—the ignition of the very first stars, and the formation of galaxies from these stars along with their central BHs—via a process that still remains completely unknown. This epoch of reionization, or "cosmic dawn," is presently at the very frontier of astronomical research, with the detection of starlight from a few $z > 7$ galaxies recently detected by Hubble and Spitzer. It is thought that these first galaxies form at local overdensities in the dark matter distribution by the accretion of almost pure hydrogen/helium gas, cooling through lines in the far-IR to sub-mm bands and finally forming stars or "seed" BHs, either from zero-metallicity Pop III stars (~100 $M_\odot$) or via some form of direct collapse (~10^4 $M_\odot$). The lack of heavy elements ($Z/Z_{sol} \leq 10^{-3}$) during the first billion years ($z > 6$) appears essential to the formation of both Pop III stars and BH seeds. The growth of these seeds proceeds either by (gravitational-wave-emitting) mergers and/or by accretion. The ESA mission LISA will "hear" SMBH mergers out to cosmic dawn at low masses, providing a complementary multi-messenger insight into this process.

Studying the first light epoch sets the bar for new key capabilities in the X-ray and FIR, well beyond what is available today, and even beyond all the planned missions for the next decade (Figure J.4). These are (1) the ultra-deep (10^{-19} ergs cm^{-2} s^{-1}) X-ray sensitivity limit over a large FOV (~1 sq. deg in ~25 Msec); and (2) the ability to undertake FIR broadband, highly multiplexed, spectroscopic surveys. Two of the Probe concepts, the Advanced X-Ray Imaging Satellite (AXIS) and the Galaxy Evolution Probe (GEP), were proposed to pursue some of this science using similar technological approaches to Lynx and Origins, respectively, albeit in smaller implementations scaled to fit within the $1 billion cost cap for Probe missions. Neither meets the required capabilities outlined above: AXIS has only one-third the collecting area of Lynx, and it is proposed as a 5-year mission. Its deep survey is 2–5 times less sensitive than that proposed for Lynx, missing the required sensitivity target by factors of 3–4 (depending on exposure times). AXIS would fail, therefore, to detect BH seeds at $z > 6$ and discern subtle features in

the X-ray luminosity function that are essential to this science.[13] *GEP* incorporates a smaller mirror than Origins, which reduces its sensitivity by a factor of 9, putting high-redshift ($z > 2$) galaxies and AGN out of reach. More importantly, its mapping speed is down by a factor of 50 compared to Origins, and the long wavelength cutoff adopted for its spectrometer (193 µm) means that important diagnostic lines cannot be detected for redshifts greater than 2.5, and its spectral resolution (R=200) is much too low to velocity-resolve the gas flowing into and out of galaxies.

In summary, a Flagship program is clearly required to enable the cosmic dance science. The panel envisions reoptimized versions of Lynx and Origins that preserve the essential capabilities, which for the purposes of this report are called Fire and Smoke, respectively. Both missions would be developed together as a single program, and launched contemporaneously, with a common science team to enable evaluation of trades both within and between them. While optimized for studying cosmic dance science, Fire and Smoke would also enable a broad range of high-priority science in other fields (as illustrated in Table J.3).

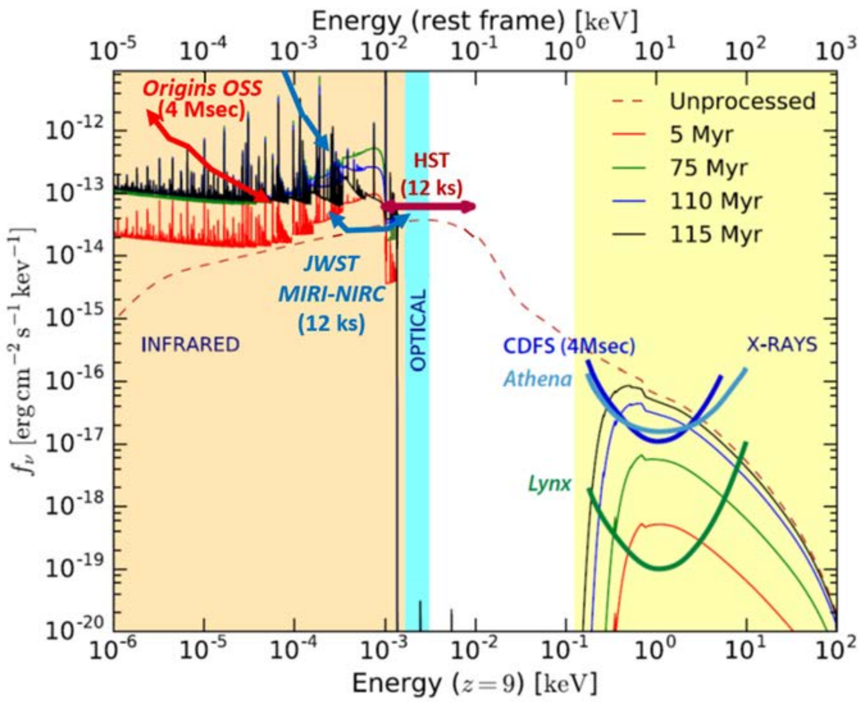


FIGURE J.4 FIR to X-ray spectral energy distribution for a black hole seed at z = 9. The Origins sensitivity curve is the thick red line with arrows on the upper left, the Lynx sensitivity curve is the thick green line on the lower right. Lynx detects all stages of black hole growth from 5 Myrs (thin, solid red line) after accretion begins, while Origins detects the later stages (> 75 Myr, thin green line). The Origins OSS sensitivity plotted is binned to a resolving power (RP) of 3. The [OIV] 25.9 um line is detectable by Origins at the native resolving power of 300 and is an important diagnostic of black hole mass and accretion rates. SOURCE: Adapted from F. Pacucci et al., 2019, Detecting the birth of supermassive black holes formed from heavy seeds, *Bulletin of the AAS*, 51(3). Retrieved from https://baas.aas.org/pub/2020n3i117. Reproduced with permission.

[13] A. Ricarte and P. Natarajan, 2018, The observational signatures of supermassive black hole seeds, *Monthly Notices of the Royal Astronomical Society*, 481(3):3278–3292.

J.4.1 Achieving the Vision

The Flagship Program of Fire and Smoke is scientifically compelling. Successful realization of its scientific promise will require careful formulation and management in a cost-constrained environment. Both the total cost and the funding profile are important, and related, as inability to meet a planned funding profile results in stretching of the schedule and increased cost. Based on the analyses performed in conjunction with the TRACE process, the execution of an Astro2020 Flagship at the $10 billion level requires funding authorization of at least $750 million for several years.

In the assessment of this panel, a cost cap and annual funding peak in these ranges would enable launches of both Fire and Smoke by 2040. To do so, however, will require a different approach to the execution of Flagships than has generally been followed. The design and development of new capabilities and the execution of the program must be undertaken from beginning to end in a constrained environment. The panel envisions a single program office, with full integration of the science and engineering teams to make hard choices and trades, and to ensure that science per dollar is always maximized and program constraints are maintained. As the design is concurrently matured, special attention needs to be paid to identifying where performance could gracefully degrade and still allow scientific advancement, versus where potential technology or performance limitations would cause step-function degradation in observational sensitivities. With the critical science and key observations required to address that science identified, the program would proceed to advance the requisite technologies, including manufacturability, scalability, and industrialization as appropriate. As emphasized in the 2017 report, *Powering Science*,[14] providing sufficient funding up front to enable such studies is the best way to retire risks prior to mission confirmation. Maintaining cost and schedule caps for the Pre-Phase A and Phase A periods, the program would then proceed into execution (Phases B–D) only if the results of the technology development and mission planning demonstrate a viable program, inclusive of sufficient cost and technical reserves.

With the exception of the period of capability development, this is the approach now followed for competed missions that from inception to completion live in a constrained environment. That has been shown to achieve the desired effect—maximizing science per dollar while remaining within constraints. Applying the same philosophy to an initial period of capability development (technology, scalability, and manufacturability) would extend this well-demonstrated method to Flagships. There are indeed a few such success stories at this scale. AXAF (Chandra) was restructured to cost, and it came within a few percent of that cost 7 years later. SIRTF (Spitzer) is a more complete example, as it was reset as a cost-constrained mission, completely redesigned to maximize science in a constrained environment, and succeeded.

To ensure the feasibility of this approach for Fire and Smoke, the panel examined the data received through the TRACE analysis of Lynx and Origins, as well as the reports for AXIS and GEP. Recognizing that this panel cannot and should not design the cosmic dance program, the top-level cost scaling exercise described below was performed to ensure the feasibility of executing both Fire and Smoke concurrently, and within the available NASA budget.

Lower Bound. Although they are inadequate as proposed to fully explore the cosmic dance, the AXIS and GEP Probe concepts can be used to derive a reasonable lower bound for Fire and Smoke. The proposed costing for these missions, at ~$1 billion each as described in the concept reports and white papers, also assumed all requisite technology has achieved ~TRL 5 or 6 prior to mission start. Addressing the necessary technology development to achieve cosmic dance science by implementing the ~$1.1 billion technology development program defined for Lynx and Origins, and recognizing these Probe-class programs would still need to be scaled up somewhat following the incorporation of this new technology, this panel found that a reasonable lower bound for the total cosmic dance program cost is on the order of a few billion dollars. While this lower bound on the cosmic dance program cost might conceivably overlap

[14] National Academies of Sciences, Engineering, and Medicine, 2017, *Powering Science: NASA's Large Strategic Science Missions*, The National Academies Press, Washington, DC.

the cost of AXIS and GEP combined, the broader scope of the science addressed, the need for coordinated development of Fire and Smoke to make science and instrument trades, and the longer mission life all argue that this is a Flagship-class mission.

Upper Bound. The panel was able to perform a more comprehensive cost estimation for Fire and Smoke by utilizing the TRACE estimates for the individual instruments required to accomplish the cosmic dance science. For Fire, this would be a wide-field imaging detector, while for Smoke, it would be a spectroscopic survey instrument. In addition to directly using instrument costs, the remaining flight system cost estimates were scaled roughly by mass for the slightly smaller implementations that would be acceptable given the necessary sensitivity limits quoted above. Costs were scaled in this way based on discussions with the Aerospace Corporation, which indicated that mass is the primary driver in the cost models incorporated in the TRACE process. The main difference between the TRACE analysis of Lynx and Origins and the cost estimates contained within the program reports was the addition of reserves and uncertainty based on prior program actuals, representing an increase in potential cost. In this costing exercise for a potential Fire and Smoke implementation, the same wrap rates for reserves and uncertainty recommended in the TRACE reports were maintained. The panel estimates that Fire could be executed for \$4.9 billion, and Smoke could be executed for \$4.5 billion, inclusive of technology development.

Because this process maintained full margin and uncertainty, this panel considers \$9.4 billion to be a reasonable 70th percentile upper bound for this two-mission program. Further efficiencies in program cost and execution can likely be gained from additional optimization of the joint mission design in a cost constrained environment, and from potential international contributions.

Notional time-phasing of the costs from this analysis are displayed in Figure J.5. Assuming Pre-Phase A technology development can begin in 2022, launch of both Fire and Smoke could be as early as 2038, ensuring the desired contemporaneous operations of the two missions required to address the cosmic dance science.

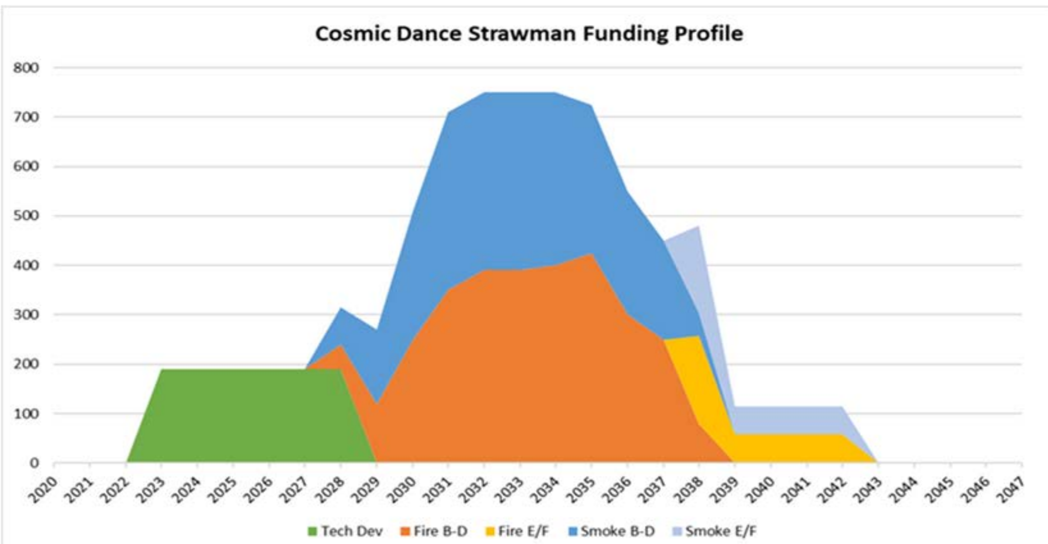


FIGURE J.5 Notional cost profiles (\$FY 2020) for the various elements of the Cosmic Dance Flagship Program, assuming a 2022 start to the technology development. Note that the peak-year funding level is consistent with a \$750 million per year funding cap in \$FY 2020.

J.5 PROBE-CLASS MISSIONS

The current NASA Astrophysics program includes missions ranging from multi-billion-dollar Flagships, to Medium-Class Explorers (MIDEX) up to ~\$300 million, to Small Explorers (SMEX) up to

~$150 million, to Missions of Opportunity (MoOs) up to $75 million, to SmallSats around $20 million to $35 million, to the recently announced Astrophysics Pioneers up to $20 million, and last, suborbital balloon flights and CubeSats in the $1 million to $10 million range. This mission suite includes a striking "hole" in the cost range from $300 million to several billion. That hole can be filled with the proposed Probe class. Several groundbreaking missions have fallen in this cost range in the past (e.g., Compton, COBE, RXTE, Fermi, Spitzer, Kepler, with Compton and Spitzer being two of the four Great Observatories).

Twelve white papers describing concepts for Probes were submitted to EOS2, encompassing the full range of the electromagnetic spectrum, as well as other messengers. They collectively demonstrate the great potential of Probe-class missions. Table J.2 briefly lists the Probe concepts that the EOS2 panel reviewed. Almost all of the science topics they cover have been highlighted in the Science Panel Key Questions. Assessing these inputs, it was evident to the panel that a wide range of strong mission concepts, including some that would achieve greater than order of magnitude leaps in performance, would likely be submitted in response to a new Probe competition.

TABLE J.2 EOS2-Related Probe-Scale Mission Concepts

Mission Concept	Lead Author	Closest Predecessor	Science Capabilities	Spectral Coverage
FARSIDE	Burns	N/A	z > 10 neutral hydrogen and SETI search on lunar far side; exoplanets; heliophysics	200 kHz–40 MHz
PICO	Hanany	Planck	CMB polarization anisotropy	21–799 GHz
CMB Spectral Distortions	Kogut	FIRAS	CMB spectral distortions	10–6000 GHz
GEP	Glenn	Spitzer, Herschel	Star formation and SMBH growth over cosmic time	400–10 µm
TSO	Grindlay	N/A	UV–mid-IR time domain astronomy follow up	5.0–0.3 µm
AXIS	Mushotzky	Chandra, Athena	Growth and fueling of SMBHs; transient universe; galaxy formation and evolution	0.3–10 keV
STROBE-X	Ray	RXTE	Compact objects; X-ray counterparts; time domain astronomy	0.2–50 keV
HEX-P	Madsen	NuSTAR	Accreting compact objects; extreme environments around black holes; neutron stars	2–200 keV
TAP	Camp	Swift	Time-domain astrophysics	0.4 keV–1 MeV
AMEGO	McEnery	Compton, Fermi	Multi-messenger; γ-ray studies of neutron star mergers; supernovae; flaring AGN	200 keV–10 GeV
POEMMA	Olinto	N/A	Ultra high-energy cosmic rays and cosmic neutrinos from space	Cosmic rays > 2×10^{19} eV Neutrinos > 20 PeV
MFB	Michelson	N/A	Fills gaps in frequency coverage between LIGO and LISA	Gravitational waves 10 mHz–1 Hz

Institution of a Probe class of missions would enable a broader NASA Astrophysics program, more balanced in size, cost, wavelength, and messenger coverage, that would better address the extraordinary range and richness of 21st century astrophysics. A truly open competition would be most responsive to the community and would enable new opportunities. However, in the view of the panel, a more limited competition, focused on particular strategic areas, may be appropriate for the first Probes, given some pressing priorities in the field. Such a strategic competition approach has been invoked in the past for Planetary Science New Frontiers missions. Below, two high-priority areas for strategic Probe competitions are identified: Time Domain and Multi-Messenger Astrophysics, and Early Universe Cosmology.

J.5.1 Time-Domain and Multi-Messenger Astrophysics

In several of the science panels, time-domain astrophysics emerged as a key scientific priority for the next decade (see Table J.3, especially the entries for compact objects and energetic phenomena). With the first LIGO/VIRGO detections of gravitational wave events, and new exciting results from cosmic ray and neutrino detectors, it is now clear that astronomy is no longer restricted to the electromagnetic spectrum; it is the era of multi-messenger astrophysics. The current modest rate of transient events will increase dramatically in the near future with planned advancements to LIGO, and with the onset of operations of the Vera C. Rubin Observatory (former LSST), that will generate 10 million time-domain alerts per night. Sifting through that data stream to identify and follow up the most exciting transients will be a major challenge to the existing system of observatories, at all wavelengths.

At present, the U.S. “workhorse” observatory for space-based time-domain studies is the Neil Gehrels Swift Observatory. Swift is an aging medium-size Explorer mission. It was launched in 2004, and although it has no expendables, its future longevity is uncertain. Outside the United States, a few other time-domain missions (see Table J.2) are now under development, notably the Gravitational Wave High-Energy Electromagnetic Counterpart All-Sky Monitor (GECAM), a Chinese mission (launched December 2020), and the Space Variable Objects Monitor (SVOM), a French-Chinese mission (2021 launch). The former combines two spacecraft 180 degrees apart, thus providing true 100 percent sky coverage in the 6 keV–5 MeV band. GECAM will immediately distribute 1-degree localizations of gamma-ray bursts (GRBs) to the community for follow up observations. SVOM, on the other hand, will be equipped with soft X-ray, gamma-ray, and optical instruments that will enable follow-up on board the mission. It will also promptly distribute GRB positions together with their main prompt gamma-ray properties, and the magnitude of their early afterglows. However, neither mission will have significant, if any, U.S. involvement, and the pair of them will not address all of the needs of the future time-domain program. While these two missions will certainly contribute to the study of time-domain phenomena in the universe, they will not adequately serve the U.S. community that has pioneered this field over the past few decades. Therefore, new NASA-led time-domain missions with enhanced capabilities are urgently needed, both to ensure long-term continuity in this developing core field and to successfully capitalize on the science that will come from advanced gravitational wave detectors and the Rubin Observatory.

Space-based platforms provide access to those bands that are undetectable from the ground: gamma-rays, X-rays, ultraviolet, and the mid- to far-IR. Historically, these bands have proven crucial to transient event detection, as well as event characterization and classification. A future system needs to include the following features: (1) detection capability at X-ray/gamma-ray energies with near 4π sr coverage; (2) prompt event localization at the few arcsecond level or better; (3) rapid-slewing for follow-up imaging and spectroscopy at X-ray, ultraviolet, and IR wavelengths; (4) long-term monitoring in these same bands; and (5) a data system capable of issuing fast alerts to the community with all essential information.

Rather than advocate for a specific mission in this field, the panel suggests instead that NASA create a coordinated strategic program in time-domain astrophysics that provides the capabilities described above, potentially capitalizing on the international missions that are operational. These could be achievable with a mix of different implementations: either a single Probe mission or a suite of medium- and small-size experiments that address some of the requirements. It is essential, however, that all of those capabilities are simultaneously available in space, so coordination in the mission developments and launches is of paramount importance. A competitive selection of the mission architecture based on targeted NASA research announcements that explicitly call for proposals to meet these objectives would harness the ability of the broad community to devise creative solutions to achieving the science at the lowest cost. That is a new paradigm for NASA Astrophysics, but the panel believes it is required to meet the science needs of this field.

An appropriate total life-cycle cost of this coordinated program would be up to $1 billion over the decade. This estimate comes from the costing analyses of specific Probe-class mission studies that meet

these science goals, as well as from the panel members' own knowledge of various Explorer-class proposals.

During the course of the EOS2 panel's deliberations, a time-domain astrophysics working group was established within Astro2020 to raise and address issues that are common to multiple program panels. The working group identified a suite of policy suggestions for this area of research to be most productive. Of particular interest to EOS2 are (1) the need for an open data policy for time-domain alerts on both NASA and NSF public facilities; (2) a suggestion that future NASA research announcements include requirements for time-domain capabilities; (3) an explicit recognition of the importance of simultaneous operation of multiple missions and facilities broadly covering multi-messenger astrophysics; and (4) consideration of the fact that time-domain investigations typically involve large collaborations, so that large author lists do not adversely affect the careers of promising young scientists in this field.

J.5.2 Early Universe Cosmology and Fundamental Physics

As detailed by the Panel on Cosmology, tremendous progress has been and continues to be made on observational and experimental data to study profound and fundamental questions about the universe on the grandest scale. The results have led to a simple empirical "concordance cosmological model" that unifies a wide range of cosmological phenomena, agreeing well with observational results that have improved by orders of magnitude over the past two decades. These results, however, do not obscure the fact that the key ingredients of that model—inflation, dark matter, neutrinos with nonzero mass, and dark energy—are not naturally explained by the "standard model of particle physics," which has been equally successful at accounting for the properties of particle collisions at high energy accelerator facilities. The age-old quest to understand the Universe on the grand scale is far from over.

To address the major science questions identified by the Panel on Cosmology, the cosmic microwave background (CMB) remains the single most important phenomenon that can be observed (see Table J.3). The CMB is the oldest light in the universe, emitted 13.8 billion years ago from an expanding spherical surface now 45 billion light years in radius. It is the cold, 2.7 K afterglow of the Big Bang. The CMB is a direct probe of physical conditions in the early universe, 370,000 years after the Big Bang. It is also a backlight to everything else observed in the intervening space and time. An enormous amount has already been learned from CMB measurements, and there is much that remains to be learned on topics ranging from fundamental physics to the formation and evolution of galaxies and clusters of galaxies, as highlighted as a Discovery Area by the Panel on Galaxies. At large angular scales, there is the prospect of detecting and characterizing relic gravitational waves from the Big Bang through their effect on CMB polarization. This has major implications for cosmology, since it provides insight into a critical phase when the infant Universe expanded by a factor of $\sim 10^{26}$ in 10^{-32} s. It is also important for fundamental physics as it gives a new handle on particle interactions at energies forever unattainable in terrestrial laboratories. At smaller angular scales, precision measurements of temperature and polarization anisotropies of the CMB will determine the sum of neutrino masses, map the location of the dark matter in the Universe, find tens of thousands of galaxy clusters out to the highest redshifts, and, in combination with observations of galaxies at shorter wavelengths, (e.g., by DESI, Rubin Observatory, Euclid, the Roman Space Observatory, and SPHEREx) illuminate the evolution of the entire universe over cosmic time. The absolute temperature spectrum of the CMB also contains unique information about the universe.

The history of the CMB field is that of continuously improving ground and sub-orbital experiments, punctuated by comprehensive measurements from satellite missions (COBE, WMAP, and Planck). This trend will continue into the next decade. CMB observations will ultimately be limited by the accuracy with which emission from all other astronomical sources in the universe can be separated out, and by systematic errors. Ground and space measurements are complementary. From the ground, CMB polarization observations offer better angular resolution, using large aperture telescopes that would

be expensive to fly in space, combined with excellent performance in a handful of atmospheric windows. Space, with its unrestricted access to all frequencies over the whole sky, its capability for uniquely accurate absolute calibration using the orbital motion of the spacecraft around the barycenter of the Solar system, and its freedom from atmospheric and other interfering signals, offers the lowest systematic errors and foreground residuals. Important progress in the next decade can be made by ambitious observations from the ground on angular scales of roughly ten degrees and smaller. Space observations will unquestionably be needed for the best foreground separation and lowest systematic errors on all angular scales, and especially on angular scales of greater than about ten degrees.

A future dedicated CMB space mission, with a goal of measuring the polarized signal to the fundamental limits set by foregrounds, would realize these goals. This will require substantial early work to (1) improve all aspects of the detector systems (e.g., mm-wave filtering and coupling, CMB-noise limited detectors, low-noise stable readouts, low-noise cryogenics); (2) enhance the ability to simulate and separate foreground emission from the CMB; (3) simulate and mitigate the effects of systematic errors; and (4) use this knowledge to optimize the design of a flight system. This preparatory work could begin immediately, eventually resulting in a Probe-class mission, with a final design based on a targeted competition, selected near the end of the decade. The appropriate investment in technology development in advance of that selection is estimated by the panel to be ~$100 million, spent over the decade.

J.6 PROGRAM BALANCE

Science return, technological advancement, and effectiveness in training the next generation of researchers are maximized by a program comprising missions of many sizes, ranging from large Flagships down to the smallest scale where useful results can be obtained. Such a program is referred to as "balanced," and the importance of balance has been strongly endorsed by previous decadal reviews. Recent developments have extended the small end of the range of space missions down to SmallSats, Pioneers, and CubeSats. The current NASA program is working well, although some improvements can be made. As emphasized in Section J.5, the addition of a Probe class will make the program even stronger. Below, the panel comments on the other classes of mission opportunities.

Flagship Missions: Major advances in scientific capability will continue to require multi-billion-dollar missions. While such Flagship-class missions are scientifically compelling, maintaining a well-balanced portfolio requires that they be more accurately estimated and more tightly conceived and managed in a constrained environment. Past Flagship missions have focused on maximizing broad science return within a given wavelength band. Programs driven instead by a prioritized set of key science questions may make it easier to optimize design trades that control cost growth. Advances in technology designed to address such specific questions are also likely to enable unanticipated discoveries in other areas.

Explorers: Explorers have produced an incredible track record of exciting science, despite their modest size and cost. Explorers provide flexibility in the overall program not accessible to larger missions. While more narrowly focused in the scope of questions they can address, they allow for compressed time scales of development, permitting timely and rapid response to newly arising scientific questions, exploitation of the most recent innovations in instrumentation, and the opportunity for scientists and engineers to experience the end-to-end design and production of space missions. The EOS2 panel endorses a vigorous nontargeted Explorer program, maintaining a cadence of two new MIDEX, two new SMEX, and at least four MoO astrophysics missions per decade. A program budget adjustment to RY dollars is advisable.

SmallSats, Pioneers, and CubeSats: SmallSats, the new Pioneers, and CubeSats offer opportunities in space at even smaller scales, which astronomy is just beginning to exploit. Important science is achievable with these platforms when angular resolution or collecting area are not the driving requirements. At present, the selection criteria for these small missions mirror those for larger missions, including the requirement for important and new science results. The panel's view is that these missions

could take on added value and utility if the selection criteria were changed to allow for and encourage demonstration of new technology, as opposed to new science results, when such demonstrations would enable important science in subsequent larger missions and could not be done efficiently from the ground.

Suborbital Programs: Suborbital experiments including sounding rockets have played a significant role in testing technologies and training researchers, leading to better space missions. The suborbital program offers flexibility and rapid scientific return, and supports technology development and testing of new concepts, particularly in the area of detectors. MeV gamma-ray missions, for example, are an exciting area that could take advantage of the capabilities of future suborbital programs. The number of successful balloon flights has dropped over the past years. Regaining and maintaining a high tempo of balloon launches is necessary to preserve the utility of this important suborbital platform.

On-Orbit Servicing: Determination of whether on-orbit servicing is advantageous must be made carefully. Servicing requirements applied broadly can increase size, mass, schedule, and cost. Attempts to limit the extent of servicing with probabilistic analyses to identify likely failure candidates have had mixed success, as failures are generally not predictable. Missions advocated by this panel involving cryogenic components present special concerns for servicing, given the complex interfaces between cooled instruments and optics, and the need to maintain precise alignment. The panel identified three elements of successful servicing missions for consideration on future projects: provision for an attaching fixture to enable future servicing if needed; easy access for replenishing consumables, such as refueling; and detailed early planning for ground integration and test, as easy-to-make interfaces aid ground activity flow and ensure accessibility for servicing in space.

J.7 INTERNATIONAL PROGRAMS

The EOS2 panel was asked to evaluate NASA's commitment to two large projects planned for development by ESA: Athena and LISA.

J.7.1 Athena

Athena is planned for launch in the early 2030s. It will address questions on the hot and energetic universe (e.g., the origin and evolution of large-scale structures, the physical processes that govern the growth of black holes, gamma-ray bursts, and more). Athena has two instruments: the X-ray Integral Field Unit (X-IFU; 0.2–12 keV), a cryogenic X-ray high-resolution spectrometer (2.5 eV at <7 keV and R = 2800 at E > 7 keV); and the Wide Field Imager (WFI; 0.2–15 keV). *X-IFU* provides spatially resolved spectroscopy with 5" pixels over a 5 arcmin field of view with 10 μs timing resolution. The WFI provides deep wide-field (40' × 40") X-ray spectral) imaging with a pixel size of 130 μm × 130 μm (2.2" × 2.2" pixels) with energy resolution of <170 eV at 7 keV. The full FOV can be read out in <5 μs, and a fast detector readout mode can obtain 80 μs time resolution to accommodate observations of the brightest X-ray sources. The mission will be placed in a large halo orbit at L2 with a baseline mission lifetime of 4 years.

Athena is currently in an advanced stage—the configuration of its instruments is frozen, and it is currently scheduled for adoption in 2022. There is a non-exchange-of-funds agreement between NASA and ESA, with two main U.S. contributions, the detectors for X-IFU and the usage of the NASA/MSFC X-ray facility for testing the Athena mirrors. ESA will provide the spacecraft, ground segment, X-ray mirrors, SIM and service modules, launcher, and operations. The member states, in particular Germany (MPE) and France (CNES), will provide the WFI and the X-IFU, with contributions from the Netherlands (SRON), Japan (JAXA), and Spain.

The panel finds Athena to be a compelling mission with an excellent return on investment for the U.S. contributions. The EOS2 panel strongly endorses this NASA commitment in its current form.

J.7.2 LISA

LISA will be the first gravitational-wave interferometric detector in space, with an expected launch date in the early 2030s. LISA will open up the millihertz-frequency band of gravitational waves, a band rich with sources ranging from white-dwarf binaries in the Milky Way to massive black holes throughout the entire universe. U.S. participation in LISA remains a high scientific priority. NASA's planned level of contribution to both the LISA hardware and the analysis of the data are judged to be appropriate at present, although the detailed technical elements are still being worked out. It is important for the U.S. science community to play an active and prominent role in the scientific analysis of the data. The EOS2 panel strongly endorses the NASA commitment to LISA in its current form.

J.8 CONCLUSIONS

The next decade promises to be an extremely exciting period in the history of space astronomy, with the launches of JWST, the Roman Space Observatory, and other smaller missions. Many fields of astrophysics have reached a mature state, where attention is now focused on answering long-term key questions, rather than simply searching for new phenomena. It is clear to the EOS2 panel that advances are most likely to come from a truly panchromatic approach, one that makes full use of the unique opportunity provided by platforms in space—access to the entire electromagnetic spectrum, free of obscuration, distortion, and background contamination by Earth's atmosphere. Opportunities for discovery space will also remain strong with an emphasis on panchromaticity.

The panel was presented with a wide array of white papers presenting a plethora of exciting new ideas for future space experiments. The proposed Flagship missions, Lynx and Origins, were subjected to particular scrutiny. While each of these mission concepts is tremendously promising in its own right, the panel judged that the contemporaneous presence of both an advanced X-ray mission and an advanced far IR mission would be even more compelling, given the strong complementarity between X-ray and FIR observations. The panel suggested that Lynx and Origins could be jointly reoptimized to yield two smaller, but achievable missions, Fire and Smoke, that would together execute a single program focused on studying the cosmic dance between BHs and their host galaxies as the Universe evolved.

The panel strongly endorses the proposal for a new Probe-class of missions filling the hole between Explorers and Flagships. While a fully competitive Probe class would harness the creativity of the community, the panel endorses two areas for new Probe science that can be highlighted for strategic competition: Time Domain and Multi-Messenger Astrophysics, and Early Universe Cosmology. These fields are ripe for discovery, address fundamental questions, and the requisite technology is either mature, or near-ready with modest additional investment.

Last, the panel strongly reendorses the concept of "balance" within the NASA program. Smaller mission opportunities are essential to train the next generation of space astronomers, test out new technologies, and address specialized scientific questions.

If this broad vision is realized, the vast majority of high-priority science questions and discovery areas highlighted by the science panel reports will be addressed and answered. Table J.3 illustrates the coupling between these questions and the mission concepts highlighted above, with brief comments on which specific new capabilities are key to addressing them.

TABLE J.3 Science Panel Questions and Discovery Areas Mapped to the EOS2 Vision

Science Panel Question or Discovery Area	Which EOS2 Mission Addresses the Question or Discovery Area **D**—Science for which the mission is specifically designed. **S**, **G**—Additional science to which the mission can make a Strong or Good contribution, but for which the mission is not specifically designed.
Panel on Compact Objects and Energetic Phenomena	
1. What are the mass and spin distributions of neutron stars and stellar black holes?	**S, Fire[a]:** Continuum shapes of disk components, broad Fe-K line shapes. **D, TDA[b]:** Continuum shapes of disk components, broad Fe-K line shapes.
2. What powers the diversity of explosive phenomena across the electromagnetic spectrum?	**S, Fire and Smoke[c]:** Late-time transients monitoring for many types. **D, TDA:** γ-, X-ray, IR discovery rapid follow-up of transients.
3. Why do some compact objects eject material in nearly light-speed jets, and what is that material made of?	**S, Fire:** Imaging of galactic and extragalactic jets. **D, TDA:** γ-, X-ray monitoring of jet spectral evolution.
4. What seeds supermassive black holes, and how do they grow?	**D, Fire and Smoke:** Detecting and characterizing the population of seed black holes in the low-metallicity era at $z > 8$.
DA. Transforming our view of the universe by combining new information from light, particles, and gravitational waves	**S, Fire and Smoke:** Late-time behavior of GW events. **D, TDA**: γ-, X-ray, IR rapid follow-up and characterization of GW events; monitoring neutrino sources such as blazars. Detect short GRBs from jetted BH-BH mergers and NS-BH mergers in conjunction with GW observatories.
Panel on Cosmology	
1. What set the Hot Big Bang in motion?	**D, CMB[d]:** Detection of primordial gravitational waves would significantly narrow models of the early universe and provide strong support for inflation.
2. What are the properties of dark matter and the dark sector?	**S, Fire**: Study of clusters of galaxies (with Roman Observatory, other lensing); annihilation line searches. **D, CMB:** Lensing of CMB.
3. What physics drives the expansion and large-scale evolution of the Universe?	**D, CMB**: Measurements can test the change of expansion rate and equation of state over cosmic time.
4. How will measurements of gravitational waves reshape our cosmological view?	**G, Fire:** Gravitational wave counterparts. **D, TDA:** Gravitational wave counterparts. **D, CMB**: Through primordial gravitational waves, the CMB probes fundamental physics at energy scales unattainable on Earth and thereby informs how the universe began.
DA. The Dark Ages as a cosmological probe	—
Panel on Exoplanets, Astrobiology and Solar System	
1. What is the range of planetary system architectures, and is the configuration of the solar system common?	—
2. What are the properties of individual planets, and which processes lead to their diversity?	—
3. How do habitable environments arise and evolve within the context of their planetary systems?	**S, Fire:** Planet atmosphere loss rates owing to host star irradiation, winds. Coronal activity of planet-hosting stars, populations. **S, TDA:** Observe flaring from magnetically active M-dwarfs to construct first full record of X-ray heating of exoplanet atmospheres.
4. How can signs of life be identified and interpreted in the context of their planetary environments?	—
DA. The search for life on exoplanets	—
Panel on Galaxies	
1. How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?	**D, Fire and Smoke:** Study of $z > 6$ star formation and black hole growth. **D, TDA:** Detect long GRBs at $z > 6$ to probe star formation rate in epoch of reionization. High-z cutoff of long GRBs to probe era of population III stars. **S, CMB:** Reionization optical depth and redshift.

2. How do gas, metals, and dust flow into, through, and out of galaxies?	**D, Fire and Smoke:** Outflows driven by AGN and stars versus cosmic time using Fe-K, H_2, OH absorption. **D, Smoke:** Metallicity and dust properties of galaxies as a function of cosmic time from z = 8 to z = 0 via IR lines and dust features.
3. How do supermassive black holes form, and how is their growth coupled to the evolution of their host galaxies?	**D, Fire:** Measuring luminosity function of rapidly growing black holes in the first Gyr. **D, Smoke:** Measure star formation and black hole accretion rates since cosmic dawn. X-ray and IR light penetrates even heavily dust-obscured regions. **S, TDA:** Extreme hard X-ray flaring of highly beamed blazars at high z to probe formation and evolution of SMBHs.
4. How do the histories of galaxies and their dark matter halos shape their observable properties?	**S, Fire and Smoke:** Measuring the properties of both the hot gas filling the dark matter potential wells and the cold matter into which it cools to form stars as a function of environment and galaxy properties.
DA. Mapping the circumgalactic medium and the intergalactic medium in emission	**S, Fire and Smoke:** Imaging and mapping CGM, IGM, and ICM metallicities as a function of environment and cosmic epoch, resolving out field sources. **S, CMB:** SZ studies of galaxy clusters.
Panel on the Interstellar Medium and Star and Planet Formation	
1. How do star-forming structures form, evolve, and interact with the diffuse ISM?	**S, Smoke:** Map star-forming regions. Measure fine-structure emission lines from O, C, Ne, S, N, Fe, Ar, and Si to determine dynamics, physical conditions, and mass in different ISM phases. **S, Fire:** Measuring X-ray emission from young protostars in star-forming regions.
2. What regulates the structure and motions within molecular clouds?	**S, Smoke:** Cloud energetics and structure mapped with O, C, and N fine-structure lines and molecular lines including mid- to high-*J* CO rotational lines, H_2O, OH, and NH_3
3. How does gas flow from parsec scales down to protostars and their disks?	**S, Fire:** Effect of protostar activity on the surrounding disk. **S, Smoke:** Measure H_2O emission lines tracing ice and water vapor from 10 K to 1000 K.
4. Is planet formation fast or slow?	**S, Smoke:** Velocity-resolved tomography reviews the accretion rates of O, H_2O, H_2, and HD.
DA. Detecting and characterizing forming planets	—
Panel on Stars, the Sun, and Stellar Populations	
1. What are the most extreme stars and stellar populations?	**G, Fire:** Magnetic field structures from coronal properties; coronal activity of populations.
2. How does multiplicity affect how a star lives and dies?	**G, Fire:** Coronal activity; image “jets” of Miras.
3. What would stars look like if we could view them like we do the Sun?	**G, Fire:** Coronal activity versus rotation; asymmetries in SN explosions.
4. How do the Sun and other stars create space weather?	**S, Fire:** Trace coronal mass ejections from flaring stars, especially magnetically active M dwarfs and AGB stars. Detect, image supernovae, supernova remnants, determine expansion rates **S, Smoke:** Measure mass outflows and ISM enrichment from supernovae.
DA. Industrial-scale spectroscopy	—

K
Report of the Panel on Optical and Infrared Observations from the Ground

EXECUTIVE SUMMARY

We face a future filled with extraordinary opportunities. New ground-based Optical-Infra-Red (OIR) observational facilities are central to addressing the most pressing and fundamental questions in astronomy and astrophysics, as assessed by the six Science Panels of the Astro2020 decadal survey.[1] The importance of some of these questions transcends the boundaries of science: How did we get here? Are we alone? To exploit these opportunities, the United States—which has led the world for decades in ground-based OIR astronomy—must overcome some enormous challenges.

First, to maintain a leadership position in the 2030s and beyond, investments of an unprecedented scale by NSF in ground-based OIR astronomy will be required. The panel has carefully evaluated the proposal to Astro2020 to create and fund a unified U.S. Extremely Large Telescope Program (U.S.-ELTP) that will combine the resources and capabilities of NSF's NOIRLab, the Giant Magellan Telescope (GMT), and the Thirty Meter Telescope (TMT). Combined with Key Science Programs facilitated by NSF's National Optical-Infrared Research Laboratory (NOIRLab), this U.S.-ELTP will create a system for the broad U.S. community that is fully competitive with, and complementary to, the European Extremely Large Telescope (E-ELT), and one that maximizes synergies with the current U.S. multi-billion-dollar bi-hemispheric system of ground-based astronomical facilities. The programmatic challenges facing the U.S.-ELTP are daunting indeed, and it is not at all clear that there will be adequate financial capacity to complete the construction and fund the operations of this two-telescope system. The panel has reached the consensus that the rewards of a successful outcome are high enough for NSF and the other GMT and TMT partners to be given the opportunity to try to achieve this success. The panel believe a zero-ELT outcome will gravely damage the U.S. astronomy community for decades.

The second challenge is posed by the need to exploit the immense investment that has already been made in the past 30 years to create a powerful and flexible U.S. ground-based OIR system. This is particularly pressing because the scientific payout from a U.S.-ELTP is over a decade away. The panel has reviewed fifty thoughtful white papers from the ground-based OIR community, and has identified a set of thematic areas in which relatively modest investments (e.g., at or below the level of the NSF Mid-scale Research Infrastructure-2 (MSRI-2) program) in existing telescopes could reap major returns during the 2020s. The panel also highlights several opportunities in this medium-scale range to build new special-purpose telescopes or telescope arrays. In some cases, these could be interagency projects (e.g., NSF and NASA, or NSF and DOE). In other cases, they could be in the context of an international partnership. Last, the panel emphasizes the importance of modest strategic investments in technology development and software, and in the further development of the systems-level approach to optimizing the performance of the OIR system in an era of time-domain/multi-messenger astrophysics.

For this plan to succeed, there needs to be a fundamental change in the way in which the federal, state, and private funding sources for ground-based OIR facilities interact. A broad partnership is

[1] See Appendix A for the overall Astro2020 statement of task, for the set of panel descriptions that define the panels' tasks, and for additional instructions given to the panels by the steering committee.

necessary for the United States to maintain leadership. NSF will need a major boost in the Major Research Equipment and Facilities Construction Funding (MREFC) line, a robust MSRI program, and a new model for how operations of new facilities are paid for. If we can accomplish all of this, we can fully reap the extraordinary scientific harvests for decades to come.

K.1 INTRODUCTION

K.1.1 Setting the Stage

It is axiomatic that major scientific discoveries are driven by new technology, and in no field is this clearer than in astronomy. Galileo Galilei did not invent the telescope, but he was the first to use it to observe the sky and record his discoveries. His book *The Starry Messenger* (1610) reported on his observations of the Moon, Jupiter, and the Milky Way. These observations revolutionized our understanding of the cosmos and ushered in centuries of discoveries to come based on ever-more-powerful telescopes. The era of astrophysics can be said to have begun roughly a century ago, launched by the construction of large telescopes armed with spectrographs. For many decades, the United States was the unrivaled leader in the construction and utilization of such facilities, from the 100" at Mount Wilson (1917) to the 200" at Mount Palomar (1948). This U.S. leadership was made possible largely through an unmatched level of philanthropy.

Such days are over. Starting in the late 1960s the level of funding provided by the U.S. federal government and of other nations produced a suite of telescopes that matched the capabilities of the largest private/state-funded facilities. This situation has continued into the current era of very large telescopes, which started to come into operation in 1990s. The importance of the next generation Extremely Large Telescopes (ELTs) has been recognized for at least 20 years, and indeed an ELT was the top ground-based recommendation of the 2000 decadal survey. Yet, as we survey the landscape today, we see that the scale of investment needed to construct and operate the next generation of Extremely Large Telescopes (ELTs) is severely straining the financial model that has served the U.S. astronomical community so well for over a century (which has largely segregated private/state-, and federally funded telescopes). Initially two competing ELT projects with major U.S. involvement emerged: the Giant Magellan Telescope (GMT) and the Thirty Meter Telescope (TMT). Since the previous decadal survey, *New Worlds, New Horizons* (NWNH), there has been an enormous amount of work done to retire the most challenging technical risks to the construction of GMT and TMT. These two projects have now joined forces with NSF's NOIRLab to propose a unified U.S.-ELT program seeking substantial federal funding for, and providing access to both ELTs by the U.S. community.

In addition, two major new ground-based OIR telescopes have just seen (or soon will see) first light. The 4 meter Daniel K. Inouye Solar Telescope is the largest solar telescope in the world, with a focus on understanding the Sun's explosive behavior. The Vera C. Rubin Observatory (the highest ranked ground-based project in NWNH) will undertake an astronomical survey, the Legacy Survey of Space and Time (LSST) that will fully open the window of time-domain astronomy. Together, these two facilities cost just over $1 billion in 2020 dollars for construction, and continuing investment will be needed to realize a commensurate return on investment.

The past decade has also seen investment in new capabilities for existing U.S. ground-based OIR telescopes. The twin national 4-meter telescopes have been converted to wide-field survey machines with the Dark Energy Camera (Blanco), and the Dark Energy Spectroscopic Instrument (Mayall) made possible by an NSF/DOE partnership. The 3.5-meter WIYN telescope is now equipped with a state-of-art spectrograph (NEID) for precision radial velocity measurements of exoplanets, thanks to an NSF/NASA partnership. The development and deployment of multi-object and integral-field spectrographs, and of advanced Adaptive Optics and High Contrast Imaging systems has made these technologies powerful

tools on the largest telescopes. Pioneering work on optical/near-IR interferometers has opened up a whole new terrain mapped with high angular resolution.

In this report, the panel assesses the investments needed in the next decade to ensure that the United States remains at the forefront in addressing the most important and enduring scientific questions (many of which can only be answered with the next generation of ground-based OIR observatories—see Section K.2 below). In the case of the ELTs (discussed in Section K.3), the cost will be considerable, with the investment needed now. The scientific payoff—while immense—will be more than a decade away. The panel argues in Section K.4 that additional investment is essential so that, in the meantime, we can fully realize the potential of the powerful and flexible OIR system constructed at great cost over the past few decades.

K.1.2 Input to the Panel

The analyses reported below are based on extensive input to the panel. The panel individually discussed 50 white papers from the community. These were used primarily to define the topical areas described in Section K.4 and guided the panel's thinking in these areas. The panel had two presentations by NSF: one in which they provided an overall budgetary framework and entreated the panel to think boldly, and a second in which they clarified the nature of the cross-divisional support in NSF for solar astronomy and the role of Astro2020 in providing guidance in this area.

The panel received and analyzed the more detailed "Request for Information Version 1" (RFI 1's) from the GMT, TMT, and NSF's NOIRLab for the U.S.-ELT Program, CHARA (OIR Interferometer), COSMO (solar observatory), and the Mauna Kea Spectroscopic Explorer (massively multiplexed spectrograph). The panel received and analyzed the less detailed RFI 2's from the FOBOS and MegaMapper massively multiplexed spectrographs, the NPOI and MROI OIR interferometers, and the Liger and WFAO next-generation AO systems. In the case of the U.S.-ELT program, the panel received full Technical, Risk, and Cost Evaluation (TRACE) analyses. (See Appendix O.) The panel went through three iterations of questions and answers with the U.S.-ELT Program based on the RFI 1's and initial TRACE results. The panel also had a face-to-face meeting with the U.S.-ELT Program leads and key personnel (from AURA, GMT, NOIRLab, and TMT). The panel did not request a TRACE for the other projects with RFI 1 or 2, because we did not deem them as being high-priority candidates for NSF MREFC-funding.

K.2 THE SCIENCE FRONTIER

Before considering the different proposed programmatic elements in the future U.S. Ground-Based OIR system, the panel would like to set the stage by highlighting four areas at the science frontier for which these facilities would play the most essential role. These are drawn from the reports of the science panels, representing this panel's synthesis of their analyses.

K.2.1 Exoplanets and Astrobiology

Is the Solar System a cosmic rarity or a galactic commonplace? How do Earth-like planets form, and what determines whether they are habitable? Is there life on other worlds? The National Academies *Exoplanet Science Strategy* and the Panel on Exoplanets, Astrobiology, and the Solar System have identified two ground-based capabilities in optical and infrared astronomy that are essential to realize the great opportunity in exoplanets and astrobiology that is open before us:

Firstly, the GMT and TMT will open an unprecedented discovery space in the study of planet formation, mature gas giants, and even terrestrial worlds. The unprecedented *contrast and angular*

resolution of the GMT and TMT will enable profound advances in imaging and spectroscopy of entire planetary systems, over a wide range of stellar and planetary masses, semimajor axes, and wavelengths, including both reflected light and thermal emission. For low-mass stars, the reduced glare of the central star may permit the GMT and TMT to image temperate rocky worlds. These facilities will also allow the detection of newly formed planets in their natal disks, providing the ground truth for the time scale of planet formation and permitting studies of the dynamical interaction between disks and planets. With the high spatial resolution of the GMT and TMT, researchers will finally be able to search the inner parts of planet-forming systems. The unprecedented *light gathering capability* of the GMT and TMT, coupled with high-resolution optical and infrared spectrographs will be powerful tools for studying the atmospheres of transiting and nontransiting close-in planets. For the closest and least massive stars, these observatories may detect molecular oxygen in the atmospheres of transiting terrestrial planets orbiting within the habitable zone of their stars, landmark discoveries with implications far beyond astronomy and even science.

Secondly, as mass is the most fundamental property of a planet, the astronomical community needs to develop the ability to measure the radial velocities of stars to a precision sufficient to measure the masses of their attendant temperate terrestrial exoplanets. These mass measurements are essential both for determining the planet's bulk composition (and, by inference, its formation history), and for interpreting studies of the planetary atmosphere, since the atmospheric scale height depends on a combination of surface gravity, temperature, and mean molecular weight. Radial velocity measurements are currently limited by variations in the stellar photosphere, instrumental stability and calibration, and spectral contamination from telluric lines. Hence progress will require a coordinated national initiative in extreme precision radial velocities that includes observers, instrument builders, stellar astrophysicists, heliophysicists, and statisticians. Although the bulk of this effort will involve smaller ground-based telescopes, the photon gathering capability of the GMT and TMT will also play a role, enabling photon-noise-limited Doppler precision of several cm/s on time scales of a few minutes, allowing researchers to disentangle the signals from the stellar photosphere and orbital motion.

The grand opportunities now afforded by exoplanets cannot be addressed with a single telescope: The optimal targets for some of the most demanding exoplanet studies will be rare and demand access to the entire celestial sphere. In addition, the technology roadmap to enable the full exoplanet science potential of GMT and TMT will be realized only by leveraging the existing network of U.S. centers and laboratories and current 8- to 10-meter class facilities.

K.2.2 The Fundamental Physics of the Universe

Over the past century some of the most revolutionary discoveries in fundamental physics have come from OIR astronomy: from the confirmation of Einstein's prediction of the trajectory of light passing close to the sun and Hubble's discovery of the expansion of the universe, to the discovery of dark matter and of an accelerating universe (likely owing to dark energy). In this vein, there are a host of exciting paths to follow over the next decade and beyond. Here the panel focuses on cases in which ground-based OIR observations are essential.

Tests of General Relativity (GR): The Galactic Center is an ideal laboratory for testing GR in the context of a supermassive Black Hole (BH). In the coming decade, it will be possible to use the next generation of Adaptive Optics to precisely measure the orbits of stars close to Sag A*. Such measurements will test both the Einstein equivalence principle and the form of the Kerr metric around a BH, allowing a search for a massive graviton or additional interactions (like scalar fields). By using ELTs to go 5 magnitudes fainter than the confusion limit today, there will likely be multiple stars with periods as short as 1 to 2 years, providing even more stringent tests of GR. With the improved sample and astrometric and radial velocity precision enabled by ELTs, it will be possible to detect GR effects that have no analog in classical dynamics such as the precession of the periapse and the Lense-Thirring or frame dragging effect, which is owing to the spin of the BH.

The nature of dark matter. Models of dark matter span an enormous range in particle mass, and in the degree to which the dark matter particles self-interact. The combination of light-gathering power and high angular resolution will enable the new ELTs to test these models. By measuring precise radial velocities and proper motions, it will be possible to construct maps of the 3D stellar orbits at dwarf galaxy centers to map the radial distribution of dark matter, to thereby determine whether self-interacting dark matter is required. Similarly, precise maps of gravitationally lensed background sources utilizing the sensitivity and spatial resolution of an ELT will make it possible to determine the dark halo mass function at low masses, which depends on the nature of dark matter.

Tests of the expansion history of the universe will not only constrain *the nature of dark energy*, but can also look for signs of "*new physics*," which is hinted at by the so-called "H_0 tension" (an apparent difference in the Hubble constant based on local scales versus the cosmic microwave background). Ground-based OIR astronomy will play a crucial role in this. Key programs include improved measurements of the evolution of the large-scale structure of the universe (made using the next generation of massively multiplexed spectrographs), and improved and refined measurement of the distance scale using new standard candles, standard sirens (electromagnetic counterparts to gravitational wave events), and standard clocks (time delay measurements of lensed Type Ia supernovae). These techniques really require the sensitivity and spatial resolution provided by ELTs to achieve their full potential. ELTs are required to reach 1 percent precision H_0 measurements with multiple techniques. With ELTs, tip of the red giant branch distances will reach deep into the Hubble flow to ~100 Mpc, enabling large numbers of direct measurements, while geometric eclipsing binary distances will be possible for all Local Group galaxies anchoring these distances. Gravitational lensing systematics, both the mass distribution of quasar hosts and the line-of-sight mass distribution, can be minimized with deep and high-resolution IFU observations. Standard siren host galaxies are likely to be faint; ELTs will be needed for spectroscopic follow-up to obtain redshifts.

Understanding the physics of cosmic inflation. Maps of the large-scale structure of the universe made possible with galaxy redshift surveys undertaken with a future generation of massive-multiplexed wide-field OIR spectrographs are crucial here. For example, a clear detection of non-Gaussianity in the primordial fluctuations would rule out the simple "single field" inflation scenario. The detection of departures from a power-law form of the primordial power spectrum would offer the opportunity to explore the history of the primordial seeds of large-scale structure during their production in the era of inflation. These require sampling significantly larger volumes with a much larger number of galaxies than can be done with existing or nearly complete facilities.

K.2.3 Galaxy Evolution

In the next decade and beyond, ground-based OIR facilities will remain key to unraveling the fundamental processes that drive galaxy evolution. Future capabilities will transform our view of galaxies, their constituent components, and their vital link with the circum- and intergalactic medium (CGM/IGM) via baryon cycling. Specific areas of progress are captured by the themes identified by the Science Panel on Galaxies (1) the evolution of the IGM and the sources of radiation during the first billion years from cosmic dawn throughout the epoch of reionization; (2) the gas, metals, and dust flows into, through, and out of galaxies at all epochs; (3) the formation of supermassive black holes and the coupling of their growth with that of their host galaxies; (4) the impact of the assembly history of galaxies and dark matter (DM) halos on their observable properties; and (5) the discovery and investigation of the first galaxies.

The complex interplay between baryons and DM, and the richness of physical processes involved, call for multi-scale multi-wavelength approaches probing the full ecosystems across time. *Ground-based OIR systems play a pivotal role* by accessing a wealth of key spectral tracers of stars, gas, and metals. Transformative advances rely on progress along two synergistic axes: spatially+spectrally detailed information and multiplexing. Key enabling capabilities are (1) R~5000–50000 spectroscopy; (2)

resolution in the infrared down to 0.01–0.02” (roughly 100 pc); (3) dense/full spectroscopic coverage up to ~2 arcmin scales (roughly 1 Mpc at z > 1); and (4) 10–100x higher sensitivity than existing facilities.

Detailed information is vital to unveil physical processes at the level of individual stars and their surrounding nebulae in the Local Group, of individual 100pc-scale star-forming complexes and globular cluster progenitors at z > 1, and of the first galaxies at z > 10 with compact <<1 kpc sizes. The characterization of their dynamical, chemical, and thermal state requires velocity resolution <10–100 km/s. Achieving the corresponding angular resolution of 10–20 mas and spectral resolution of several 10^3 to a few 10^4, with the necessary sensitivity for such faint sources, *will only be possible with ground-based ELTs, coupled with high-performance AO systems.* With 5–10 times sharper views, ELTs+AO will map z > 1 galaxies in similar detail as we have now for galaxies a mere 100 Mpc away—a drastic advance reaching fundamental sub-galactic components. The ELTs will thus uniquely enable major breakthroughs in galaxy evolution.

Multiplexing is essential to build up sample statistics and to cover wide contiguous areas. This is needed for (1) a complete census and detailed characterization of ~10^8 stars and the low-mass satellites that encode the Milky Way’s chemo-dynamical history and the baryon-DM interplay, (2) to connect directly individual galaxies and their surrounding CGM/IGM, and (3) to map the distribution and properties of the first generations of galaxies holding keys to the history and topology of reionization. Ongoing projects and future concepts for massively multiplexed spectrographs—be they multi-object or integral field instruments—will deliver the necessary 10-100 times larger samples. Coupled with high sensitivity and angular resolution, they will reach unexplored low mass/luminosity regimes that are very sensitive to the physics of star and galaxy formation, overcome crowding in nearby studies, and greatly boost observational efficiency for distant sources.

K.2.4 A Stellar Renaissance

Over the past 15 years, the solar and stellar astrophysics community has ignited a scientific renaissance through a remarkable investment in facilities for global synoptic and high-resolution solar observations, extreme-precision stellar radial velocity, and milliarcsecond imaging of nearby stars, as well as ultra-widefield surveys focused on high-precision position measurements, high-cadence time-domain measurements, medium-resolution optical-infrared spectroscopy, and multi-band precision photometry. Ground-based OIR facilities have played, and will continue to play, the central role.

A crowning achievement for ground-based solar astronomy is the recent completion of NSF’s 4 m Daniel K. Inouye Solar Telescope (DKIST), whose unprecedented spatial resolution and spectropolarimetric sensitivity will open a new window on the magnetohydrodynamic phenomena that affect convective motions and drive the storage and release of magnetic energy (reconnection). These in turn result in sunspots, flares and eruptive events, solar energetic particles, coronal heating, and perhaps most fundamentally the solar cycle, all of which have impact on our home planet through space weather. Solar measurements made from the ground (DKIST, GONG, SOLIS, GST, and other observatories, as well as the planned COSMO, ngGONG, and DKIST-II) and from space (SDO, SOHO, Solar Orbiter (with ESA), Parker Solar Probe, and the in-development Solar C/EUVST (with JAXA)), the coming decade will provide significant new insight into physical processes that apply to all stars. High time-cadence observations by NASA’s Kepler mission have opened an analogous window on the interiors of distant stars via asteroseismology and photometry, while high spatial-resolution observations with ground-based optical interferometers such as CHARA and VLTI are producing astonishing insights into the surface phenomena and asymmetries of nearby stars.

Meanwhile, the combination of ground- and space-based ultra-wide-field imaging, spectroscopic, and high-precision astrometric surveys (e.g., SDSS-IV Apogee, ESA’s Gaia, DOE/NSF Dark Energy Survey) are enabling a giant leap forward in researchers’ ability to disentangle galactic structure and more precisely constrain the dynamical and chemical evolution of the Milky Way over its entire lifetime, as well as providing new, more precise measurements of fundamental stellar properties and lifecycle stages

across the HR diagram. Indeed, such large surveys have been critical in detecting examples of statistically rare, short-lived stellar types.

The renaissance in solar and stellar astrophysics will continue and accelerate in the decade ahead thanks to OIR ground-based facilities that have just started operation or will begin soon such as NSF/DKIST, DOE/NSF/VRO, DOE/NSF/DESI-Stellar, NASA/NSF/NEID and ESO/4MOST. High-time cadence information will be added for large stellar samples by NASA/TESS and ESA/PLATO. Many of the brightest stars in those samples will be observable by ground-based optical interferometers and high-spectral resolution spectrometers and spectropolarimeters to open a new window on the connection between stellar surface and interior phenomena. Looking further into the future, the spatial resolution of ELTs will isolate single stars below the main-sequence turnoff, map the orbits of tight binaries, and enable proper motion selection in distant clusters in currently inaccessible environments throughout the Local Volume. The ELTs will also be useful to follow up the detections of candidates for Pop III stars in the halo of the Milky Way and in other Local Group galaxies.

Affiliating and effectively using these highly multi-dimensional peta-scale data sets will require new approaches to managing and analyzing positional, photometric and spectroscopic information, likely incorporating database and machine-learning techniques developed outside astronomy. Astronomers have always worked closely with mechanical, optical, electrical, and computer engineers. In the decades ahead, astronomers will also embrace working with data scientists in ways not broadly appreciated even 10 years ago.

K.3 THE U.S. EXTREMELY LARGE TELESCOPE PROGRAM

K.3.1 Introduction

We stand at a watershed moment. A new generation of Extremely Large Telescopes is essential to answering the most important questions in astronomy and astrophysics in the 2030's and the decades that follow. It seems clear that without a major federal investment, the two ELT programs with U.S. stakeholders, the Giant Magellan Telescope (GMT) and the Thirty-Meter Telescope (TMT), will both fail, severely damaging U.S. astronomy for decades.

At this pivotal time, the two ELT projects have joined forces with NSF's NOIRLab to propose a radically new concept to Astro2020: a unified U.S.-ELT Program that will provide the broad U.S. astronomy community with access to both GMT and TMT in exchange for a substantial federal investment in the ELTs, and with an interface to this broad community provided by NOIRLab. The panel then considered two stark choices: Endorse an unprecedented level of NSF investment and a revolutionary new "business model" to capitalize on this investment, or cede U.S. leadership in the frontier science enabled by ELTs for decades to come. The panel argues below in favor of the first choice on the basis of the review of a proposed U.S.-ELT Program uniting GMT, TMT, and NOIRLab.

K.3.2 The Science Case

In Section K.2, the panel described four broad areas at the scientific frontier where the next generation of ground-based OIR facilities are most essential and explicitly identified there the need for ELTs, from the search for life on distant planets, to the births and lives of galaxies, to the fundamental physics of the cosmos.

Here the panel reports on its analysis of the reports of the Science Panels to identify which of the specific Questions and Discovery Areas (DAs) will need ELTs. For each of these, the panel has evaluated whether the ELTs are essential (18 cases), useful (8 cases), or not needed (4 cases) to answer the questions posed. The essential items are listed in Table K.1, which makes it clear that the science case for

ELTs is extraordinarily broad and deep (and likely unmatched by any single other new ground- or space-based observatory currently under consideration).

TABLE K.1 List of Essential Needs for ELTs in Science Frontier Panel Reports

Panel on Compact Objects and Energetic Phenomena
Q2: What powers the diversity of explosive phenomena across the electromagnetic spectrum?
Q4: What seeds supermassive black holes and how do they grow?
DA: Transforming the view of the Universe by combining new information from light, particles, and gravitational waves.
Panel on Cosmology
Q2: What are the properties of dark matter and the dark sector?
Q3: What physics drives the cosmic expansion and large-scale evolution of the universe?
Panel on Galaxies
Q1. How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?
Q2: How do gas, metals, and dust flow into, through, and out of galaxies?
Q3: How do supermassive black holes form, and how is their growth coupled to the evolution of their host.
Q4: How do the histories of galaxies and their dark matter halos shape their observable properties?
DA Mapping the circum-galactic medium and the inter-galactic medium in emission.
Panel on Exoplanets, Astrobiology, and the Solar System
Q1. What is the range of planetary system architectures?
Q2. What are the properties of individual planets, and what processes lead to planetary diversity?
Q3: How do habitable environments arise and evolve within the context of their planetary systems?
Q4. How can signs of life be identified and interpreted within the context of their planetary environments?
DA: The search for life on exoplanets.
Panel on the Interstellar Medium and Star and Planet Formation
Q3. How does gas flow from pc-scales down to proto-stars and their disks?
Q4. Is planet formation fast or slow?
DA: Detecting and characterizing forming planets.
Panel on Stars, the Sun, and Stellar Populations
The ELTs were deemed useful but not essential to the four questions and discovery area identified by the Panel on Stars, the Sun, and Stellar Populations.

K.3.3 The Components of the U.S.-ELT Program

The U.S.-ELT Program (U.S.-ELTP) as proposed to Astro2020 is made up of three elements: The GMT, the TMT, and NOIRLab.

The primary mirror of the GMT consists of seven 8.4 m segments, for a total diameter of 24.5 meters. It is a two-mirror Aplanatic Gregorian telescope yielding a 25 arcmin field-of-view (FOV) at f/8 with Ground-Layer Adaptive Optics. Alignment and phasing is achieved with Adaptive Secondary Mirrors. Laser Tomography Adaptive Optics utilizes six sodium laser beacons and edge-sensors. With this, diffraction-limited images of 10 mas at 1 μm can be achieved over a 20–30 arcsec FOV. The first generation of instruments includes the GMT Multi-object Astronomy and Cosmology Spectrograph (GMACS), the GMT Integral Field Spectrograph (GMTIFS), the GMT Consortium Large Earth Finder (G-CLEF), and the GMT Near-IR Spectrograph (GMTNIRS). The GMT will be located at the Las Campanas Observatory in Chile. The majority of the GMT partners are U.S. institutions, with

international partners in Australia, Brazil, and Korea. The site has been excavated, the site water and electrical distribution upgrades have been completed, and the sixth mirror has been cast.

The primary mirror of the TMT consists of 492 1.44 meter mirror segments, for a total diameter of 30 meters. It is a three-mirror Ritchey-Chretien telescope delivering a 20 arcmin FOV. Ground Layer AO can improve image quality over a 2 arcmin FOV. Six sodium lasers and 3 tip/tilt natural guide stars will be employed. With this Narrow Field InfraRed Adaptive Optics System (NFIRAOS), diffraction-limited images of 8 mas will be achieved over a 15–30 arcsec FOV. Instruments will be deployed at the Nasmyth platforms. The first-generation instruments include the Infrared Imaging Spectrometer (IRIS), the Wide-Field Optical Spectrometer (WFOS), and the Multi-Object Diffraction-limited High-Resolution Infrared Spectrograph (MODHIS). The TMT will either be sited at the Mauna Kea Observatory in Hawaii (MKO), or at Roque de los Muchachos Observatory in the Canary Islands (ORM). The majority of the TMT partners are international, with the participation of institutions in the United States, Canada, China, India, and Japan.

NSF's NOIRLab is the U.S. national center for ground-based, nighttime optical and infrared astronomy. It operates five programs—Cerro Tololo Inter-American Observatory (CTIO), the Community Science and Data Center (CSDC), Gemini Observatory, Kitt Peak National Observatory (KPNO) and the Vera C. Rubin Observatory (VRO). Its roles in the U.S.-ELTP will be to (1) provide access to a bi-hemispheric ELT system; (2) enable and support large-scale, systematic, collaborative research (Key Science Programs—see Section K.3.10 below); (3) provide user support, broaden participation in TMT/GMT science, and foster research inclusivity; and (4) engage and represent the whole U.S. community in GMT and TMT governance, scientific planning, and instrumentation development.

K.3.4 Technical Risks

Since *New Worlds, New Horizons*, both the GMT and TMT projects have devoted significant resources to identify and retire the major technical risks. It is the panel's assessment that—for the most part—these are now both technically mature projects which are now mastering the implementation of new technology. However, there are still issues to be solved. The TRACE analysis by the Aerospace Corporation gave both projects a medium rating in technical risk. The largest concerns for both projects were the active and adaptive optics systems.

For GMT the most important issue is the complexity of 7-segment adaptive secondary mirrors, since phasing and alignment to desired specifications across 4 modes is required. The on-axis guide star and laser guide star modes were found to be the most challenging. Quantitatively, the adaptive seven-segment secondary mirror has 675 actuators per segment. Challenges with this system could lead to schedule delays and may limit desired scientific performance across select observing modes. The Laser Tomography Adaptive Optics mode depends on six sodium lasers and edge sensing to < 15 nm RMS. Its design requires significant development.

For TMT, the biggest concern is that the NFIRAOS AO instrument is complex, with multiple sodium lasers, wave-front sensors, deformable mirrors and downstream imagers and spectrographs. It is a single point failure whose performance directly impacts two of three first-generation science instruments. It still requires significant development across multiple partners. The primary mirror integration and wave-front control system for the 492 segments involves 1476 actuators and 2772 edge-sensors, and will also be challenging. Problems with this system could lead to delays.

K.3.5 Construction Costs and Funding Model

In the panel's view, the most serious risk to the successful construction for both the GMT and the TMT projects is financial rather than technical. The cost estimates provided by the projects are considerably higher than those previously understood by the astronomy community and will severely

strain the financial capacity of even a full partnership between the projects and the National Science Foundation.

The GMT project now estimates a total construction cost of $2 billion in real year (RY) dollars. Of this amount, 20 percent has been spent to-date. An additional 10 percent has been committed by current partners. The project plans for the remaining 70 percent to come from NSF (40 percent), additional (uncommitted) funds from existing partners (15 percent), and funds from unidentified new partners (15 percent). To state it differently, even with all the funds expended and committed, and with an $800 million commitment from NSF, there is still a shortfall of roughly $600 million.

The TMT project now estimates a total construction cost of $2.65 billion in RY dollars. Of this amount, 18 percent has been spent-to-date. An additional 41 percent has been committed by the current partners. The project plans for the remaining 41 percent to come from NSF (30 percent) and additional (uncommitted) funds from current partners (11 percent). In this case, with all the committed funds and full funding from NSF, there is still a shortfall of $310 million.

The construction costs estimated in the TRACE reports are both about 20 percent higher than the project estimates ($2.4 billion for GMT and $3.1 billion for TMT). They are largely the result of conservative assumptions made by the TRACE analysis as to risk. All these numbers are summarized in Table K.2.

TABLE K.2 Summary of Construction Costs and Funds, in Million Dollars and Real Years

	GMT	**TMT**
Funds spent	$400 (20%)	$475 (18%)
Additional funds committed	$197 (10%)	$1063 (41%)
NSF ask	$800 (40%)	$800 (30%)
Missing funds	$603 (30%)	$310 (11%)
TOTAL	$2000	$2650
___________	___________	___________
TRACE delta	$400 (20%)	$450 (17%)
TRACE TOTAL	$2400	$3100
TRACE missing funds	$1003	$760

K.3.6 Programmatic Risks

The largest single programmatic risk to each project derives directly from the funding issues summarized above. *Both projects need significant additional new funding beyond the planned request from NSF. Both projects believe that the combination of the imprimatur of a top ranking in the decadal survey, followed by the full financial involvement of the U.S. federal government would make it possible to secure additional resources from existing partners and possibly from new partners.*

Before discussing the two projects, it is important to emphasize that they are based on very different funding models, each with different potential risk factors. The majority of the GMT partners are U.S. universities (plus the Carnegie Institution for Science). The model for the contribution and allocation of financial resources is strongly cash-based. In contrast, the majority of the TMT partners are international entities funded by their respective national governments. The TMT funding model is largely based on significant in-kind contributions from these partners in terms of completion of assigned technical work packages.

The GMT project's estimate of the additional funds needed in addition to the funds currently committed and the funds requested from NSF is $600 million. If the construction cost for the TRACE is adopted, this increases to $1000 million. A further risk factor for GMT is the relative immaturity of estimates for cost-to-go. Only 16 percent of these are based on signed contracts or detailed bids, with the

other 84 percent being based on “Rough-Order-of-Magnitude” or project estimates. This leads to additional uncertainty in cost and schedule.

The TMT project’s corresponding estimate of these additional new funds is $310 million. If the TRACE cost estimate is adopted this becomes $760 million. A further potential risk factor is TMT’s reliance on international partners. This is based on a model of low-cost in-kind critical components and sub-systems. While TMT has agreements in place that a given international partner would shoulder any additional costs associated with the delivery of their work packages, this model has not been tested under extreme circumstances. If this funding model is successful, the TRACE delta could be significantly decreased.

An additional potential programmatic risk for TMT is posed by the uncertainty in its choice of site. Based on the documents presented by TMT, which were analyzed by the panel and in the TRACE report, a timely decision to build TMT on ORM would not lead to an increase in cost or a delayed schedule compared to MKO. Moreover, the panel has reviewed the relevant metrics on site quality and finds that—while MKO is the superior site—the ORM site is acceptable. The largest impact would be in the thermal infrared and in the ultraviolet near the atmospheric cut-off. Despite this assessment, the choice of a site still poses a significant programmatic risk since it could adversely affect the partnership.

The TRACE analysis—based primarily on the considerations above—gave both projects a medium-high programmatic risk. The TRACE evaluation flagged the schedules as being too aggressive. GMT plans for 12 years, including LTAO commissioning, while the TRACE estimate was 13 years. TMT plans for 10 years, while the TRACE estimate was 13 years. The panel notes that there is better agreement with risk-adjusted schedules from the projects: 13.7 years for GMT and 11.2 years for TMT. Both stated their costing included the risk-adjusted schedules.

K.3.7 Life Cycle Costs

As is discussed further in Sections K.3.8 and K.3.9, the panel regards it as essential that a plan is developed to ensure that adequate funding is in place through a combination of federal and project funds to operate the U.S.-ELTs with high efficiency, and to continue to provide them with state-of-the art instruments throughout their scientifically productive lifetimes.

The two projects have developed bottom-up estimates for operations of these facilities and have budgeted for partial funding of future instruments. The panel is concerned that the estimated operations and instrumentation budgets are too lean, for both GMT and TMT. GMT has budgeted $30.6 million per year for operations and $10 million per year for new instruments. The total annual amount of $40.6 million represents about 2.1 percent of construction costs (all numbers being in $2020). TMT has an annual budget of $33.3 million for operations and $13.7 million for new instruments. The total annual budget of $47 million represents 2 percent of construction costs. In comparison, the corresponding fractions are 4 percent for the E-ELT and 5 percent for the VRO and ALMA.

It is the panel’s assessment that the currently proposed operations budgets are too low to support the type of highly efficient, flexible mode that will be needed to capitalize fully on the financial investment. Likewise, the budget for future instrumentation will not be adequate to ensure a continuing line of state-of-the-art new instruments.

K.3.8 Consideration of the U.S.-ELT Program

In this section, the panel considers the case for the proposal to Astro2020 by the U.S.-ELTP for federal investment that would unite the GMT, TMT, and NOIRLab. The panel first discusses the question of whether this should be done, and then asks whether it can be done, from a financial point-of-view.

The panel emphasizes here that a single proposal was received from the U.S.-ELTP (not separate GMT and TMT proposals). The panel has therefore considered this proposal for investment in a two-ELT

system. In Section K.3.9, the panel addresses what would happen if NSF and the two ELT projects cannot firmly commit to the total financial resources required.

K.3.8.1 The Strategic Case for the U.S.-ELTP

The panel believes that the scientific case for a U.S.-ELT program is strong and compelling. The technical risks, while challenging, are manageable. The greatest threat to these projects is financial in nature, and could potentially be retired through an NSF investment, if that is part of a full and robust financial plan with the other partners. Given the status of the GMT and TMT projects, and the on-going construction of the E-ELT, the time to invest is now.

Here, the panel summarizes the strategic considerations that informed its thinking. These considerations form the core of the argument to Congress for federal support at such a substantial level.

Ensuring U.S. Leadership: The construction of the E-ELT is well under way. Absent a federal investment, it seems clear that the U.S. community will cede leadership in what has been the backbone area of observational astrophysics for over a century. In this context, a federal investment only makes sense if the program it enables is at a level that effectively competes with and complements the E-ELT capabilities over a long lifetime of discovery.

The E-ELT (39-meter aperture) is larger than either the GMT (24.5 meters) or TMT (30 meters). A U.S. federal investment in either GMT or TMT on its own, will not achieve parity. In particular, the sizes of the professional astronomical communities in the United States and Europe are similar, so that a fraction of a single ELT would underserve the U.S. community.

There are a number of ways in which *the sum of the GMT and TMT* can offer important advantages over the E-ELT, some of which is described below. Here the panel emphasizes two points. The first is that the United States has a powerful legacy of astronomical discoveries using a bi-hemispheric OIR system of telescope. Unlike Europe, this bi-hemispheric approach has enabled the U.S. community to undertake major surveys or follow-up on high-stakes targets in any hemisphere required. The second is that GMT and TMT have respective unvignetted fields-of-view that are six and four times larger than the E-ELT in terms of solid-angle, making them more efficient for undertaking surveys. The panel notes that the U.S. community has extensive experience in large astronomical surveys. This wide-field, bi-hemispheric capability makes the U.S.-ELTP complementary to the larger-aperture E-ELT.

Capitalizing on Investments and Assets: The strongest argument in favor of a bi-hemispheric U.S.-ELTP is the importance of being able to fully exploit the synergy made possible by the investments already made in bi-hemispheric facilities. No matter how powerful the ELTs may be, they are part of a system of ground-based facilities that have been constructed at great cost (over $3 billion). These facilities are needed to unlock the scientific power of the ELTs, and *vice versa*. These facilities are (by hemisphere, in alphabetical order):

- ***North:*** *ARC 3.5, Gemini North, HAWC, HET, Keck I and II, LBTO, Mayall (DESI), MMT, Palomar 5 m, PTF, Pan-STARRS I and II, Sloan Telescope (SDSS N), Subaru (HSC and PFS), SMA, JVLA, WIYN (NEID)*
- ***South:*** *ALMA, Auger Observatory, Blanco (DECam), DuPont (SDSS-S), Gemini S, Magellan I and II, SALT, SOAR, VRO, SPT (CMB/SZ), IceCube*

In addition to these ground-based facilities, space-based facilities are inherently all-sky, and require bi-hemispheric ground-based facilities to fully exploit their power. Last, there are fields on the sky that are only accessible to either a northern- or southern-hemisphere ELT, and represent either an enormous existing investment in space-based and/or ground-based observations, or are astronomically unique, including:

- ***North:*** *EGS, Euclid NEP, GOODS-N, Kepler, M 31, SDSS (main), TESS-N*

- ***South:*** *GOODS-S and Euclid DF-S, Euclid DF-Fornax, LMC, SMC, TESS-S, Roman ST HLS, SDSS (S)*

There are also very rare and important sources to be discovered that may only be visible from one hemisphere, and there are monitoring programs where the longitudinal coverage afforded by two ELTs would be valuable.

Risk Mitigation: There are serious programmatic risks for each of the ELTs, leading to multiple ways in which either could fail. By pursuing the as-proposed U.S.-ELTP, the risk of a zero-ELT outcome is reduced.

A Platform for Innovation: A two-telescope system provides a broader platform for innovation. Firstly, two telescopes provide more nights of observations to the community, making a broader scientific program possible. Secondly, while the first generation of instruments on GMT and TMT have overlapping capabilities, this need not be the case in the future. The U.S.-ELTP could decide on a strategic plan through which future generations of instruments are complementary and access is shared across the partners. This would add to the advantage that the smaller plate scales for GMT and TMT mean that instruments will be less expensive compared to those on the E-ELT, making it possible to achieve greater diversity in capabilities for a fixed investment.

The NOIRLab as a "Force-Multiplier": A crucial strategic consideration in the panel's view is the use of the NOIRLab as a "force-multiplier" in terms of return to the community. The panel agree with their plan (described in Section K.3.10 below) to develop a set of large strategic community-driven Key Science Programs on a scale that would greatly boost the impact of the U.S.-ELTP. In fact, the panel believes that the true power of this approach really requires that the non-U.S.-federal partners also contribute observing time to these programs and are fully engaged with the broader U.S. community in terms of selecting, designing, and executing these programs.

A Trans-Pacific Partnership and a Worldwide ELT System: The successful formation of the U.S.-ELTP would join together the United States in a scientific partnership spanning the Pacific Ocean, from North America (Canada and the United States) and South America (Chile), to Australia, East Asia (China, Japan, Korea), and South Asia (India). This partnership would be able to forge collaborations with the European ELT from a position of strength and in so doing form a powerful worldwide ELT system that fully exploits the complementary strengths of these facilities for decades to come.

K.3.8.2 Can This Be Done?

As discussed in Section K.3.5, the intention of the U.S.-ELTP is to request a total of $1.6 billion RY dollars funded by NSF's MREFC line. These funds would be split evenly between GMT and TMT. Can this be accommodated within NSF's budget?

NSF presented an aspirational MREFC budget for the coming decade that represents a very significant increase over the mean annual budget over the past decade. The U.S.-ELT program has also presented a baseline for when the funds from NSF would be needed. The comparison between this request and the notional NSF MREFC budget is shown in Figure K.1 below.

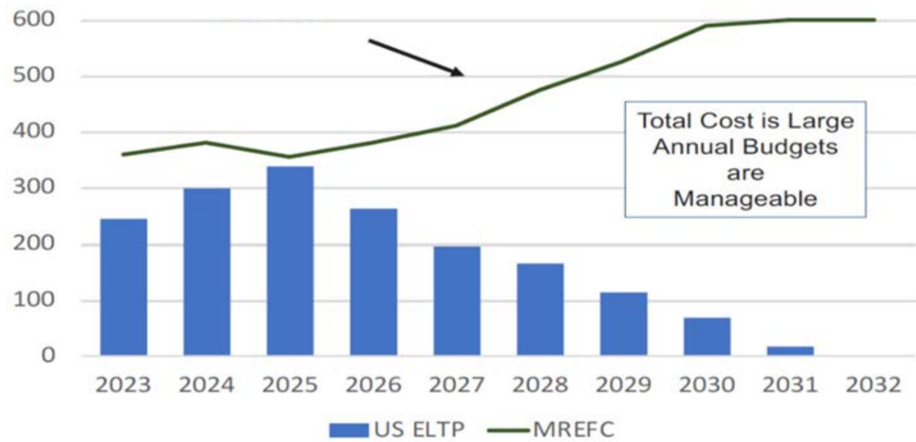


FIGURE K.1 Comparison between the U.S.-ELT request and the total notional NSF MREFC budget. The green line (marked by an arrow) represents NSF's guidance to Astro2020.

From this, two conclusions are obvious. Yes, the U.S.-ELTP can fit within the NSF guidance. However, the MREFC line pertains to the entire NSF program (not just astronomy), and over the 5 fiscal years starting in 2023, the U.S.-ELTP by itself would require 70 percent of the total MREFC budget. In other words, NSF might have the capacity to provide the requested funds, but only if the U.S.-ELTP was the number one major NSF construction project until late in the decade.

The panel has expressed its concerns in Section K.3.7 that the proposed funding level for operations and new instruments may be insufficient to realize a commensurate return on the enormous investment in construction. Even the NSF share of these budgets as proposed by the projects is problematic, since the current NSF model is one in which operations costs of observatories built using MREFC funds are borne by the Division for Astronomical Sciences from its existing budget. This cannot work in this context. The panel was briefed by NSF on this issue and told that changes are being considered. Nevertheless, this remains a major concern of the panel.

K.3.9 Conditions for an NSF Investment

The panel believes that maximum flexibility on the part of NSF, in the terms of the final arrangements it negotiates with the U.S.-ELTP, will lead to the best outcome. However, in order for NSF to agree to provide the requested funds to the U.S.-ELTP, the panel believes that there are critical conditions to be met. These are of two kinds. The first are programmatic in nature, and the panel assumes that they would be met through the due diligence of NSF. The second concerns the critical role played by the NOIRLab on behalf of the U.S. community.

K.3.9.1 Programmatic Conditions

The following programmatic considerations would be preconditions for NSF investment in the U.S.-ELTP: First, the partners in GMT and TMT present a full financial plan that has all the resources to complete construction, including adequate contingency. Second, NSF caps their investment at the requested amounts of $800 million (RY) for each project. Third, the share of observing time secured by NSF is in direct proportion to their investment in fixed-year dollars (and hence larger than the notional 25 percent if $1.6 billion is allocated). Last, the current U.S.-ELT partners and NSF need a firm and realistic plan for life-cycle costs for at least the first decade (covering both operations and new instruments). This plan would be rigorously reviewed by NSF to determine whether the proposed budgets represent the best

trade-off between cost and scientific return. The NSF portion of the operations budget would explicitly include a program of competitively selected grants to enable research using data obtained as part of the Key Science Programs (Section K.3.10) undertaken with the ELTs.

If it is not possible for all parties to commit to the full financial requirements, NSF could then examine whether it is possible to achieve a full financial commitment for one telescope, and decide whether to proceed with that plan to realize a U.S.-ELTP involving NOIRLab.

K.3.9.2 The Role of the NOIRLab

The second set of conditions deals with the role that the NOIRLab would play in making the U.S. community a full scientific partner in the U.S.-ELT program and ensuring that the return on the federal investment to this community is maximized. The U.S. federal government, through the NSF investment, will be far-and-away the largest single stakeholder in each of these telescopes. Thus, a strong U.S. leadership position would be enabled by empowering NOIRLab to act on behalf of the full U.S. community.

The federal investment in the U.S.-ELTP needs to accomplish more than simply providing much of the missing financial resources needed to construct and operate these observatories. This investment can be leveraged as part of the negotiations to create the right framework: the NSF resources can be used in part to provide the “glue” to form a true partnership. NSF can ensure this by empowering the NOIRLab to play a central role in the partnership. More specifically, the panel envisages that the NOIRLab would be charged to:

1. Provide access to a bi-hemispheric U.S.-ELT system.
2. Engage and represent the whole U.S. community in GMT and TMT governance and scientific planning.
3. Provide user support, broaden participation in GMT and TMT science, and foster research inclusivity.
4. Assist the community in developing Key Science Programs (see Section K.3.10 below), and solicit the active involvement of the entire GMT/TMT U.S. and international partnership in these, including contribution of their observing time.
5. Develop the capability for queue scheduling applied to all GMT/TMT observations to optimize the way in which data are taken and maximize their immediate and legacy scientific value.
6. Facilitate plans for new instruments (including a process of open competition for building these instruments by members of the U.S. community when federal funds are involved).
7. Ensure that well-calibrated and well-characterized data and resulting data products are in an easily used archive containing all TMT/GMT data.

K.3.10 The Key Science Programs and the Need for U.S.-ELT Science Centers

The panel believes that the Key Science Programs are a critical component of the U.S.-ELTP. These are community-driven ELT programs that are being facilitated with the help of the NOIRLab. They are based on the principle of creating a powerful scientific legacy through systematic investment in large-scale, transformative research.

These are projects on scales difficult to realize within shares of any of the current GMT/TMT partners, and are envisaged as utilizing at least half of the U.S.-ELTP observing time. As emphasized above, the panel believes the true value of the programs will only be realized if all the ELT partners participate in them, with not only scientific contributions, but also observing time.

These are planned to have broad, inclusive scientist participation via open collaboration models, to harness resources of a diverse research community, spread scientific benefits widely through the community, and produce data products with high archival reuse value.

While the panel finds these to be an exciting concept, it worries that there is no current funding mechanism in the United States that can support them, nor are they part of the current U.S.-ELTP operating budget as presented to the panel. Nonetheless they are critical to supporting the development of the most effective science program from the U.S.-ELTP. Funds for Science Centers would enable collaborations of people and institutions to carry out leadership science and create data, data products, and knowledge infrastructure that would enable multiple generations of scientific usage from large coherent data sets. The creation of a funding mechanism to create Science Centers may potentially be a way for NSF to partner with private foundations, a partnership of increasing interest at NSF.[2]

The national benefit of such Science Centers associated with the U.S.-ELTP Key Programs would be many-fold. They would provide the data and knowledge infrastructure necessary to carry out these large, long-term, multi-partner programs. They would produce myriad scientific discoveries derived from these highly leveraged, large coherent data sets, drive the hardware development and produce software tools for enabling science, and create a team of people capable of exciting and effective public engagement and stewardship of the major funding investments made within the U.S.-ELT partnership. The panel believes that without these Science Centers, the revolutionary potential of a U.S.-ELTP will not be fully realized.

K.3.11 Summary

The coming generation of Extremely Large Telescopes (ELTs) will be scientifically essential for decades to come. They will be necessary to address the majority of the questions posed, and discovery areas identified, by the Science Frontier Panels. The European Southern Observatory is well-along in the construction of the European ELT (E-ELT) with an aperture of 39 meters. It follows then that without a U.S. response, the United States will be ceding leadership in astronomy (and not just ground-based OIR astronomy) for at least a generation. It is also important to emphasize that—historically—large ground-based OIR telescopes remain highly productive for 50 to 70 years. The cost of the U.S.-ELTP is high, but it can be "amortized" over many decades.

There are currently two ELTs with significant U.S. private and state participation, the Giant Magellan Telescope (GMT) and the Thirty Meter Telescope (TMT). It seems clear that neither project can be successfully completed without an unprecedented level of federal support from NSF. The GMT and TMT, in concert with NSF's NOIRLab have therefore proposed to Astro2020 to create an integrated U.S.-ELT Program (U.S.-ELTP) that will provide access to both GMT and TMT in exchange for federal investment of $800 million in each of these two telescopes and also in the necessary NOIRLab infrastructure.

In the panel's view, in order to be fully competitive with the E-ELT, to ensure full synergy with the extensive system of U.S. bi-hemispheric ground-based astronomical facilities, and to strengthen the case for this level of financial support (which will take the enthusiastic support of Congress), NSF will need to work with all three U.S.-ELTP elements to create the proposed system of two telescopes, and implement a bold and visionary new model that forms a true partnership with U.S. community engagement (enabled by the NOIRLab). NSF will be by far the largest single partner in this program, and can therefore play an appropriate leadership role.

The path ahead is very difficult. There are serious risks associated with both ELTs, and several ways each could fail. The panel believes the engagement of NSF with both projects helps mitigate the risk of no U.S.-ELTP.

---

[2] See, for example, https://www.msri.org/workshops/785.

For NSF to proceed with this plan, it would need to ensure that each of the other GMT and TMT current and future partners commit the financial and technical resources to fully complete construction and provide the funds needed for operations and a line of new instrumentation. The panel believes that the two-telescope U.S.-ELTP is needed to maintain U.S. *leadership, and that it would be disastrous if no U.S.-ELTP is realized.*

K.4 A BALANCED PROGRAM FOR THE NEXT DECADE AND BEYOND

K.4.1 Introduction

As exciting and extraordinarily important as the U.S.-ELTP would be, the scientific payoff to the community is over a decade away. Moreover, while the multi-billion dollar U.S.-ELTP price tag may seem daunting, it is crucial to recognize the comparable investment made in existing OIR ground-based observatories, starting with the two Keck telescopes in the 1990s. For telescopes with U.S. stakeholders, this amount is roughly $2.4 billion in 2020 dollars (just in initial construction costs). The Daniel K. Inouye Solar Telescope) (which has just seen first-light), and the Vera Rubin Observatory (operations starting in 2023) together account for about $1.0 billion of this total.

At great cost, we have built a powerful and flexible system of ground-based OIR facilities. How do we exploit this investment during the next decade and beyond? In this section of report, the panel describes what it believes to be the most essential components of this system. The primary goal is to highlight how relatively modest investment either in, or in support of, the existing (or soon-to-exist) ground-based OIR telescopes will pay major dividends during the decade to come. In some cases, the panel also describes possible new telescopes or telescope arrays that could either plausibly be fit into the NSF Mid-Scale Research-2 (MSRI-2) program or funded by NASA. The panel does not consider any of these facilities as rising to the level requiring the MREFC. These are described in alphabetical order.

K.4.2 Adaptive Optics/High-Contrast Imaging

Adaptive optics (AO) is one of the transformational technologies that have driven breakthroughs in many areas of astronomy and astrophysics in the past two decades, by enabling ground-based telescopes to image at their diffraction-limit. AO development has led, notably, to the discovery of the supermassive black hole at the Galactic Center, the first images and spectra of exoplanets, and direct evidence for the existence of dark subhalos as predicted by Cold Dark Matter models. AO is a key component for the success of DKIST.

Many current and new instruments at existing OIR facilities are equipped with AO capabilities. AO is intrinsically embedded in the design and operations of ELTs. Thus, the central role of AO instrumentation and the importance of further development are rapidly growing, with novel concepts pushing toward wider areas, higher performance, and extended wavelength coverage.

- Wide-field AO (GLAO, MCAO, MOAO, LTAO) delivers uniform wavefront correction over large areas, achieved by sensing the atmospheric turbulence profile with multiple laser beams assisted with natural guide stars, and serving a very wide range of areas from Galactic to extragalactic science. Examples include censuses of stellar populations in the Milky Way and Local Group where AO significantly reduces the impact of crowding, and surveys of resolved morphologies and kinematics of distant galaxies whose apparent sizes are of order of the seeing disk and smaller. Wide-field AO has matured in the 2010’s, with first systems now in science operations (e.g., GeMS on Gemini-South, ARGOS at the LBT).
- Extreme AO (ExAO) lies at the opposite end of the AO parameter space and aims to provide exceptionally high performance (Strehl ratios in excess of 90 percent) over a narrow field.

Technical implementation of the ExAO concept is highly challenging. Coupled with high-contrast observations (>10^5), the primary goal is the high-priority exoplanet science case. The challenges and scientific rewards have motivated massive efforts worldwide. Pathfinders are in operation (e.g., GPI on Gemini South, SPHERE at the VLT) but significant development notably in deformable mirror technology, speed of AO systems, and wavefront control is needed in the 2020's to achieve the scientifically driven technical requirements (contrast of 10^6 or better).

- Visible AO has high potential scientific return by opening up an entire wavelength regime to high angular resolution studies. The goal is to exploit the smaller diffraction limit ($\propto\lambda/D$) of telescopes in the optical, yet both the coherence length and time decrease at shorter wavelengths ($\propto\lambda^{6/5}$) requiring wavefront sensing at high spatial and temporal frequencies that are currently technologically challenging. This is an important developing area for the 2020s–2030s.

The panel received and reviewed RFI 2's for GNAO (at Gemini-North) and LIGER (at Keck). GNAO development was initiated as part of the NSF-funded program GEMMA (Gemini in the Era of Multi-Messenger Astronomy) for a queue-operated MCAO facility that will deliver diffraction-limited imaging over a ~1.5 arcmin field at Gemini North. The ultimate goal of transforming Gemini-North into a full AO telescope will require deployment of an Adaptive Secondary Mirror (ASM) and integration of multi-laser guide star wave-front sensors (WFS) into the Acquisition and Guiding (A&G) unit. LIGER is an innovative next-generation AO-fed integral field spectrograph and imaging camera, fully funded through the Final Design Phase (until December 2020), and will take advantage of the NSF-funded KAPA (Keck All-Precision AO) system. With investments at the level of ~$15 million each for each project (equipment and labor), over timelines of 5–10 years to completion (2025 for LIGER, 2029 completion of ASM + A&G camera for GNAO), these programs are NSF MSRI-2 candidates.

The review of the Programs (and Science) White Papers by the U.S. community, along with the RFI 2 documents, led the panel to conclude that the case for continued support of development over the next decade is strong:

1. AO/High-Contrast Imaging are key enabling technologies for high science priorities identified by the Exoplanets, Stars, and Galaxies science frontiers panels;
2. They play an essential role in boosting the scientific return and efficiency of existing facilities (e.g., Gemini, Keck, Magellan, DKIST) with modest-scale investment throughout the 2020s;
3. NSF Mid-scale program opportunities have been identified (e.g., GNAO, GmagAO-X, LIGER) to nurture existing 6–10 m class telescopes.
4. Such investments in AO systems development is a key risk mitigation strategy for ELTs, whose full resolution and sensitivity potential can only be realized with AO, and which is recognized as the most important technical risk for both GMT and TMT.

K.4.3 Extreme-Precision Radial Velocities

Mass is the most fundamental property of a planet, and knowledge of a planet's mass is essential for two reasons. First, a measurement of a planet's mass is necessary to constrain its bulk density, and in turn infer something of its composition and ultimately its formation. Second, a planet's mass is a key input for interpreting spectroscopic features in its atmosphere, since the atmospheric scale height depends on the planetary surface gravity, in addition to the mean molecular weight and temperature. There is keen interest in studying terrestrial planets, including those orbiting at Earth-like insolations around Sun-like stars, motivating a push for mass measurements to the sensitivity required for such worlds.

The National Academies *Exoplanet Science Strategy* (ESS) had two key findings related to precise radial velocities, both echoed by the Panel on Exoplanets, Astrobiology, and the Solar System:[3]

> **ESS Finding:** The radial velocity method will continue to provide essential mass, orbit, and census information to support both transiting and directly imaged exoplanet science for the foreseeable future.

> **ESS Finding:** Radial velocity measurements are currently limited by variations in the stellar photosphere, instrumental stability and calibration, and spectral contamination from telluric lines. Progress will require new instruments installed on large telescopes, substantial allocations of observing time, advanced statistical methods for data analysis informed by theoretical modeling, and collaboration between observers, instrument builders, stellar astrophysicists, heliophysicists, and statisticians.

The panel advocates that together NASA and NSF address the grand challenge of achieving the precision required to measure the masses of terrestrial planets orbiting Sun-like stars, which implies a single measurement precision of 10 cm/s and control of systematics at the level of 1 cm/s. While such measurements will be done from the ground, they are inextricably linked to the scientific success of numerous current and proposed missions, namely the legacy Kepler/K2 data set, the ongoing TESS Mission, and a future direct imaging mission. For a direct imaging mission, such precise radial velocities will identify terrestrial exoplanets orbiting nearby stars, determine when they are situated at quadrature, and remove degeneracies from the interpretation of atmospheric spectra features. NASA has tackled formidable technology challenges in the past in pursuit of its scientific goals, and the panel advocates here that this same coordinated effort be brought to the EPRV challenge. The panel concurs with the recommendation from the NAS Exoplanet Science Strategy,[4] namely,

> **ESS Recommendation:** NASA and NSF should establish a strategic initiative in extremely precise radial velocities (EPRV) to develop methods and facilities for measuring the masses of temperate terrestrial planets orbiting Sun-like stars.

Following this recommendation, NASA and NSF jointly commissioned a community-based Extreme Precision Radial Velocity Working Group, which recently presented the blueprint for a strategic EPRV initiative.[5]

K.4.4 OIR Interferometers

The renaissance in stellar astrophysics in the 2010s was driven by new data and capabilities that enabled enormous advances in precision measurements of fundamental physical properties of stars: masses, radii, and luminosities. The Panel on Stars, the Sun, and Stellar Populations's first science priority is measuring fundamental stellar properties across the H-R diagram, with an emphasis on precision stellar masses and radii. Ground-based OIR interferometry is the key enabling technology, with increased polarimetric sensitivity providing significant opportunity for new discovery. The 2010s saw significant advances in OIR interferometry, with CHARA and NPOI becoming fully operational and scientifically productive. During this same time there was steady progress in construction of the ambitious

---

[3] National Academies of Sciences, Engineering, and Medicine, 2018, *Exoplanet Science Strategy*, The National Academies Press, Washington, DC.

[4] National Academies of Sciences, Engineering, and Medicine, 2018, *Exoplanet Science Strategy*, The National Academies Press, Washington, DC.

[5] See https://exoplanets.nasa.gov/exep/NNExplore/EPRV/.

MROI configurable imaging interferometer. The OIR interferometry community is growing, and has presented a unified vision of how to advance science with these three complementary facilities in the coming decade.

The panel received and analyzed an RFI-1 from CHARA and RFI-2s from each of NPOI and MROI. These provided further information on the technical and performance goals for the projects, and presented detailed budgets and schedules to achieve those goals during the 2020s. These formed an important component of the panel's discussions to assess the scale of the required investments.

The panel concludes that continued U.S. scientific and technical leadership in ground-based optical interferometry requires strategic investment during the 2020s. Based on the RFI responses, modest ($1 million to $2 million per year) investments would promote continued growth of the interferometry user community and fund new technology development efforts. Mid-scale investments in CHARA and NPOI at the high end of MSRI-2 would enable implementation of larger telescopes, longer baselines, and advanced beam combination technologies needed to deliver the greater angular resolution and photometric sensitivity required to achieve the Stars panel's science goals. A final phase of the envisioned CHARA upgrade (replacement of all existing 1 m inner-array telescopes with new AO-equipped 2 m telescopes) would require an additional mid-scale investment during the 2030s. Achieving the potential of MROI would require an MREFC-level investment to bring the full 10-telescope array into science operation by the end of the 2020s.

The full triad of U.S. OIR interferometers is deeply complementary: no one system can accomplish the science of all three by themselves. However, NSF is unlikely to have sufficient funds to support all three. Thus, the U.S. interferometry community and funding agencies would be best served by formulating a plan to realize this goal collectively instead of through internecine competition. This will strengthen U.S. leadership and make it an important component of a balanced U.S. ground-based OIR portfolio.

K.4.5 Massively Multiplexed OIR Spectrographs

There is very strong support for massively multiplexed spectroscopy across many sectors of the science community. The survey received five white papers related to the topic (SDSS-V, MegaMapper, SpecTel, DESI, and FOBOS), and this capability was highlighted in multiple science questions by the Cosmology, Compact Objects, Galaxies, ISM, and Stars panels. The survey also received an RFI 1 from Maunakea Spectroscopic Explorer (MSE) and RFI 2's from MegaMapper and FOBOS. Furthermore, numerous past reports have heavily emphasized this capability. The support stems from the strong feeling that large-format spectroscopy is required to maximize return on investment for the large-area imagers coming online in the 2020s.

Many science panels emphasized this capability. The Cosmology panel emphasized large volume large-scale structure studies and the need for photometric redshift calibration. The stars panel discussed "industrial scale spectroscopy" in the context of galactic archeology studies of dwarf galaxies, stellar streams, and the galactic halo, as well as spectroscopically exploring stars with unprecedented photometric data (e.g., from Gaia). The Panel on Galaxies considered the importance of galactic archeology as well as statistical galaxy evolution studies and IGM tomography, and the Panel on Compact Objects and Energetic Phenomena discussed transient follow-up and dynamical searches for compact objects.

Massively multiplexed spectroscopy is required to fully realize the primary science goals of the VRO, the Roman Space Telescope, Gaia, and other surveys. Happily, investment in existing 2–10 m capabilities would achieve a large fraction of the main science goals through the continuation of projects like SDSS V, DESI, U.S. access to the Prime Focus Spectrograph on Subaru beyond their nominal missions, or investment in a next generation instrument for an existing telescope, like FOBOS on Keck.

A dedicated facility would of course provide advantages over relying solely on existing infrastructure. Most glaring is the lack of high spectral resolution (R~20,000) multi-object spectrographs.

Both the Panel on Galaxies and the Panel on Stars, the Sun, and Spectral Populations emphasized the power of this modality for galactic archeology work. MSE and SpecTel presented plans to the panel for such a mode. MegaMapper offers the widest field of view, as required for any cosmological application. Both SpecTel and MSE have larger apertures, which would be powerful for galaxy evolution and cosmological applications. Last, a Southern telescope (SpecTel or MegaMapper) will have obvious benefits for follow-up in the era of the VRO.

In all cases, the United States could envision playing a significant role in these projects through a MSRI-2-level investment, which could provide up to about 20 percent of the cost of a project like MSE, SpecTel, or up to about 50 percent of MegaMapper, perhaps split with DOE. However, the time scales are such that the panel strongly encourages investment in some combination of next-generation SDSS V, DESI, and PFS through the 2020s, which could then be followed by investment in a larger dedicated facility. The panel does not believe an MREFC-level of funding is warranted in this decade.

K.4.6 Solar Physics

The Sun is the only star observable with high spatial resolution and the only star for which global synoptic observations are made continuously. First light of the 4 m DKIST telescope in early 2020 promises significant progress in the coming decade for understanding detailed physical processes in the photosphere and the low corona related to the causes of flux emergence, the dynamo that drives stellar activity cycles, the mechanisms of coronal heating and solar wind acceleration, the fundamental process of magnetic reconnection, including the triggers for sudden release of stored magnetic energy in the star's atmosphere, and the effects of stellar activity on the habitability of exoplanets around stars more or less like the Sun. DKIST first-light instruments have break-through capabilities. Plans for the next generation can begin after initial observations indicate the most important capabilities for future instrumentation.

To fully address the survey's science goals for solar and stellar astrophysics, very high-resolution observations in the restricted field-of-view of DKIST would need to be supplemented by global synoptic measurements. Two critical measurement gaps exist that require investment in the coming decade: Measurements of the magnetic field in the global corona and greatly improved synoptic observations of the solar photosphere. Such observations of the Sun are also critical to the national priority to better understand and predict space weather and its effects at Earth and elsewhere.

Present coronal observations are mostly limited to intensity/density. Magnetic measurements are required to develop better physics of the corona and for understanding energy storage and release related to explosive events. The Coronal Solar Magnetism Observatory (COSMO), a proposed upper mid-scale project, will measure the global corona using a 1.5 m refractive coronagraph. The project passed a preliminary design review in 2018 and can provide synoptic observations of the coronal magnetic field both above the limbs and for certain structures on the solar disk. The panel received and reviewed an RFI 1 from COSMO.

The current ground-based solar synoptic network has two aging components: the Global Oscillation Network Group (GONG) and Synoptic Optical Long-term Investigations of the Sun (SOLIS). The six-site GONG network has limited capability owing to modest spatial and spectral resolution, and the more capable SOLIS instrument suite has only one site that is not currently operating. Space-based observations currently fill this gap, but have finite life expectancy and measure quantitatively the magnetic field at only one height in the atmosphere. Continuous measurements of velocity, magnetic field, and intensity at multiple heights in the photosphere and above are required. Currently a conceptual design, ngGONG would be a global network (six) of highly capable solar observatories, including coronagraphs at three sites.

Program Balance: The COSMO, ngGONG, and DKIST 2nd Generation projects are expected to compete against other AST-wide projects in a more robust NSF mid-scale project line. DKIST was an MREFC project that addresses important science goals that could not be achieved in any other way. Funding DKIST operations is a burden for the facilities budget of NSF/AST, but funding for operations of

major new facilities needs to be addressed more broadly at NSF. The existing suite of U.S. solar GB-OIR observatories includes the 1.6m Goode Solar Telescope (GST, a DKIST precursor), the Mauna Loa Solar Observatory (MLSO, a COSMO precursor) and a few smaller facilities based at universities (see the National Academies *2020 Solar and Space Physics Midterm Assessment*). Continued investment in these facilities provides critical measurements enabling new science, continuity of unique long-term synoptic observations, and capabilities for instrument development. Improving programmatic balance, including adequate/increased support for science analysis was a major theme of the 2013 Solar and Space Physics decadal survey. Coordination with in situ and remote-sensing space-based instrumentation including Parker Solar Probe, Solar Orbiter, and Solar Dynamics Observatory is essential to address survey goals.

K.4.7 Technology Development: Astrophotonics

Well established technologies developed in the past decades for fiber-optics telecommunications and other commercial applications are now emerging as potentially revolutionary strategies for a new generation of astronomical instruments. With micro-electro-mechanical-systems already adopted, for example, by the JWST, the new frontier is represented by astrophotonics, which uses fibers and optical waveguides built in solid state devices to manipulate light collected by a telescope.

The current portfolio spans a wide range of devices, including multi-core fibers, photonics lanterns, photonics spectrographs, complex Bragg gratings, on-chip beam combiners, pupil remappers, and laser frequency combs. The GRAVITY instrument at VLTI and the MIRC instrument at CHARA have successfully adopted on-chip beam combiners to combine the light collected by their 4x8 m and 6x1 m telescopes, respectively. The possibility of obtaining extremely high-precision radial velocities, of the order of a 10 cm/s or better, as well as direct imaging of exoplanets—two of the main science cases for the U.S.-ELT system—may largely rely on the maturity of single-mode fibers and on-chip nulling interferometers.

Strengthening the coordination between the most active astrophotonics research groups in the United States would optimize resources and facilitate the passage from laboratory research to industrial partnership. This could be done through the creation of a distributed, multi-disciplinary Institute of Astrophotonics to coordinate the teams working in this field. The more coordinated approach adopted by Europe (Germany in particular) and Australia has led to success and leadership in this field. A few tens-of-millions of dollars of funding over the next decade would be needed to significantly advance this technology and reestablish U.S. leadership in astrophotonics. Cost-savings for translational programs in astronomical applications can derive from the availability of past fundamental research, industrial prototypes, and available technological infrastructure for hardware production.

The versatility, compactness and lightweight of photonics devices make them ideal for space science applications. NASA, therefore, may want to invest and support these developments. Increasing the coordination between NSF and NASA, as well as other stakeholders like DOE, NIST, and so on would create a coherent and synergistic program that optimizes resources and avoids duplications, taking into account the different strategic goals of the various agencies.

K.4.8 Time-Domain Astronomy and Multi-Messenger Astrophysics

In the past decade, we witnessed the birth of multi-messenger astrophysics (MMA), and the important role that ground-based OIR telescopes played in identifying and characterizing the electromagnetic counterpart to gravitational waves from a binary neutron star merger. The healthy ecosystem of optical telescopes, including NOAO-access facilities (2 m LCOGT/Faulkes, 4 m Blanco, 4.1 m SOAR, and 8m Gemini-S) enabled a prompt localization and detailed characterization of the optical/infrared kilonova light curve and its spectrum, showing the telltale signatures of the products of r-process nucleosynthesis and their subsequent radioactive decay.

Heading into the next decade, we expect a rich landscape of discoveries in time domain astronomy (TDA). These range from wide-field surveys such as the NSF/DOE VRO in the optical and new wide-field capabilities in the radio, to gravitational wave observatory networks like Advanced LIGO and KAGRA, and ground-based all-sky very high-energy neutrino and gamma-ray and detectors like IceCube and HAWC, as well as space-based gamma-ray burst detectors. Opening up these new windows of discovery will place heavy demands on the ground-based OIR ecosystem, for localization, classification, and the characterization of transients.

In many cases, a tiered approach to follow-up OIR observations of MMA and TDA discoveries will be required, engaging the full range of the OIR telescope ecosystem: from wide-field and/or rapid slewing 1–4 m aperture telescopes for candidate discovery and filtering, to 6.5–10 m aperture telescopes for spectroscopic classification, to an extension of VRO LSST operations for deep, wide-field target-of-opportunity gravitational-wave counterpart searches, and ultimately to ELTs for detailed characterization from spectroscopy and late-time light curve evolution. Similarly, the new landscape of TDA surveys, most notably VRO LSST, will yield a population of persistent, variable sources that will demand flexible telescope scheduling for time-sensitive observations.

One of the biggest challenges in this exciting era of MMA and TDA will be the efficient, selective, and prompt allocation of follow-up observing resources. Specific needs will include (1) upgrades to enable robotic telescope operations (e.g., LCOGT); (2) coordinated, dynamic telescope scheduling software (e.g., TOMS); (3) real-time communication and data analysis infrastructure (e.g., AEON); (4) automated data reduction and calibration (e.g., using Astropy); (5) machine-learning enabled classification (e.g., Antares); and (6) systems for rapid incorporation and reprioritization from initial follow-up observations taken (e.g., TreasureMap). With investment in these cyberinfrastructure tools and robotic telescope operation and scheduling capabilities, we can fully exploit the existing ecosystem of OIR telescopes to handle this new treasure trove of transients and multi-messenger discoveries in the next decade. Furthermore, by enabling the participation of private telescopes in an NSF OIR-lab follow-up network, we can further expand U.S. OIR access and coordination for effective follow-up in the MMA and TDA era.

K.4.9 The NOIRLab and the OIR System

The U.S. ground-based OIR research community supports and benefits from a remarkable array of general purpose and specialized facilities of all aperture sizes in the 2 m to 10 m range. Financial support comes from a heterogeneous collection of federal, nonfederal, private, and international sources. The scientific passion and technical ingenuity that led to these capabilities enabled U.S. global leadership for more than 100 years. Since the late 1950s, this so-called OIR System has benefited greatly from NSF financial support for major observatories, telescopes, instruments, and data systems.

The formation of NSF's Optical-Infrared Astronomy Research Laboratory (NOIRLab) is a welcome, long-awaited outcome recommended in whole or part by several national reports including the 2000 and 2010 Decadal Survey reports. The newly constituted center provides (1) critical end-to-end (photons-to-science) infrastructure for wide-field optical imaging (Blanco/DECam, VRO) and optical spectroscopy (Mayall/DESI); (2) time-domain research (public brokers, SOAR, Gemini North, Gemini South, data pipelines, etc.); (3) access, visualization, and analysis of massive object catalogs with millions to billions of objects; and (4) wide-field, high-spatial resolution imaging at the 8 m scale (Gemini North and South). On behalf of NSF, NOIRLab operates and maintains mountain-based research parks in Arizona and Chile for facilities built, deployed, and operated by university-based science collaborations. It also facilitates the long-term development and retention of human capital. Indeed, many of the key managerial and technical leaders for the DKIST, VRO, GMT and TMT projects were initially trained at one of the NOIRLab constituent parts. Last, NOIRLab is the focus for strategic community discussion and planning at NSF's behest, see Table K.3.

Beyond NOIRLab, there are several facilities that are de facto national assets, with the W.M Keck Observatory being an example. Sustaining a university-based instrumentation and technology development program in support of the OIR System is critically important to retaining and supporting a technical workforce, as well as training students for careers in all branches of science, technology, and engineering. Equally important is continued support to the community for development and maintenance of data analysis and management infrastructure, from Astropy to cloud-based systems needed to work with petascale multi-dimensional object catalogs to science-case driven software systems needed for time-domain, multi-messenger research.

TABLE K.3 Critical Leadership Roles NOIRLab Will Play in the Decade Ahead

Coordination/leadership of the U.S.-ELT Project on behalf of the NSF investment.
Efficient and effective execution of the VRO Legacy Survey of Space and Time (LSST), including provision of world-class, open access data services for the LSST data products.
Efficient coordination NOIRLab facilities and data systems with other OIR System facilities to support time-domain observation research, especially in regard to rapid follow-up of short-duration phenomena such hypernovae, massive compact binary objects mergers, SMBH tidal disruption events, etc.
Operation of unique capabilities within the OIR System, including wide-field optical imaging (DECam), wide-field multi-object spectroscopy (DESI), wide-field multi-conjugate and ground-layer AO imagers (GNAO/GNAOI, GEMS/GSAOI), and extreme precision radial velocity spectrometer (NEID).
Execution of regular community-based strategic planning exercises in coordination with its OIR System partners for the benefit of the entire U.S. community.
Contributing to the maintenance of a healthy program of technology R&D in the United States. More explicitly, NOIRLab can provide the coordination needed to support and optimally exploit the system of groups represent the U.S. excellence in design and construction of new instrumentation.

Last, just as spectrum management at radio frequencies has been a key NSF mission for many years, the time may have arrived for NSF to take a more active role in managing the use of the optical window. For decades, all major OIR observatories have faced challenges from background light generated by ground-based lights of various kinds. In the past 12 months, OIR astronomy is suddenly facing an almost existential challenge from mega-constellations of satellites in Low Earth Orbit (LEO). Federal action and support are needed immediately, or an entire scientific field may be crippled in an unrecoverable way.

K.5 CONCLUSIONS

The exciting cases presented by the Science Panels show that even in this era of panchromatic/multi-messenger astrophysics, observations with ground-based OIR telescopes will continue to be essential for addressing many of the most important scientific questions we will face in the decades to come. These telescopes are not only invaluable in their own right, they are also essential in fully unlocking the potential for discovery and understanding provided by the other windows on the universe.

The U.S. community will need to successfully deal with some extraordinary challenges if it is to maintain a position of leadership in ground-based OIR astronomy. In fact, ground-based OIR astronomy is so intimately and intricately woven into the fabric of the discipline that researchers face challenges in maintaining leadership in astronomy as a whole. These challenges are twofold: to create a U.S.-ELT Program that is fully competitive and leverages the existing bi-hemispheric investment in astronomical facilities, while at the same time, providing the resources needed to exploit the powerful suite of existing facilities in the current decade and beyond.

Meeting these challenges requires a sea-change in the way the U.S. ground-based OIR community and its federal/state/private funding sources work together. Without a vigorous partnership

between the various components of this system, the United States will not be able to remain a leader. We need to learn from the past and face the future with boldness and vision.

L
Report of the Panel on Particle Astrophysics and Gravitation

L.1 EXECUTIVE SUMMARY

Our universe is almost certainly populated with sources more wondrous and consequential than anything we have seen or even imagined. This belief is well supported by astronomical history, where surprises have been common when new capabilities are developed. It drives the core work of astronomy—that is, observation, from which all else follows—to consistently reach for greater sensitivity.

Astronomy has been revolutionized by observations in increasingly broad swaths of the electromagnetic spectrum, for example, through imaging black holes with radio interferometry, seeing the dust-enshrouded hearts of galaxies with infrared light, and revealing constellations of high-energy sources with X rays. Multi-wavelength observations have also revolutionized the understanding of physics, for example, through establishing the foundations of the Hot Big Bang cosmology, testing general relativity with binary pulsars, and revealing the cauldrons where the elements are made. The hallmark of this work, for which projects with sensitivity up to hard X-ray energies have largely been developed, funded, and carried out as part of astronomy programs, is its precision for localizing and measuring sources.

Astronomy is now also being revolutionized by observations with new messengers—gravitational waves, neutrinos, gamma rays, and cosmic rays—that greatly complement and are leveraged by observations in conventional astronomy. The hallmark of this work, for which the projects have largely been developed, funded, and carried out as part of physics programs, is its ability to probe extremes of energy, fields, and density. We characterize these four probes as new messengers owing to huge advances since the previous decadal survey. Gravitational waves and very high energy neutrinos were detected for the first time. For gamma rays, there have been dramatic advances in the energy range, angular resolution, and number of sources. For cosmic rays, there have been dramatic advances in the precision and composition of the spectra, plus hints of sources. The breakthrough discoveries enabled by these new messengers will be broadened and deepened by ongoing and planned experiments.

Astronomy can also be revolutionized by greater efforts on diversity, equity, and inclusion, which will lead to new perspectives and discoveries, as well as societal benefits.

From observations with these new messengers, we now know that our universe contains objects with surprising properties that were hidden from conventional astronomical studies. These objects include the following:

- Extreme gravitators, with incredibly strong fields that distort spacetime (including electromagnetically dark mergers of black holes);
- Extreme accelerators, with total power and per-particle energy far beyond laboratory experiments (including neutrino sources that are presently unknown and which may be hidden from electromagnetic observations), and;
- Multi-messenger sources, where some of these processes are also revealed by electromagnetic radiation, especially gamma rays (including mergers of neutron stars, gamma-ray bursts, flares of active galactic nuclei, and more).

The techniques of particle astrophysics and gravitation, in addition to probing such sources, are also essential for fundamental studies of cosmology, including inflation, dark matter, and tests of new physics.

The Particle Astrophysics and Gravitation (PAG) program panel (hereafter "the panel") was charged to "identify and suggest to the Decadal Survey committee a program of federal investment in research activities" within its topical scope. The panel reviewed observatories using these new astronomical messengers and considered technology-development and other needs to support cutting-edge programs that probe both the sources noted above and the properties of the new messengers. While the panel reviewed white papers for many worthy potential projects, it suggests that only a fraction of them are compelling for significant investments in the 2020s. This report is focused on those projects.

The panel sees a compelling opportunity to dramatically open the discovery space of astronomy through a bold, broad multi-messenger program, with three components:

- *Neutrino program:* A large-scale (MREFC) investment by the National Science Foundation (NSF) in IceCube-Gen2, to resolve the bright, hard-spectrum, TeV–PeV diffuse background discovered by IceCube into discrete sources and to make first detections at higher energies.
- *Gravitational-wave program:* Medium-scale investments in three bands (kHz, nHz, and mHz) to develop a rich observational program: Cosmic Explorer, with NSF support for technology development to set the stage for large-scale investments and huge detection rates in the 2030s; the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), with NSF support for expanded operations in the 2020s; and the Laser Interferometer Space Antenna (LISA), with National Aeronautics and Space Administration (NASA) support for a broad scope of activities to build a vibrant U.S. community for significant science contributions in the 2030s.
- *Gamma-ray program:* Medium-scale investments that support observations over a wide energy range, with two components. (In this report, for simplicity we use "gamma-ray" to mean photons at or above hard X-ray energies.) First, a NASA Probe-scale mission, targeted to multi-messenger astronomy, with sensitivity in the keV–MeV–GeV range and with capabilities for the identification, localization, and characterization of transients. This would be selected by competitive review; potential projects include the All-sky Medium Energy Gamma-ray Observatory (AMEGO), the Advanced Particle-astrophysics Telescope (APT), or the Transient Astrophysics Probe (TAP). Second, U.S. participation in TeV-range ground-based experiments for precision studies—for example, the Cherenkov Telescope Array (CTA) and the Southern Wide-Field Gamma-Ray Observatory (SWGO)—as NSF medium-scale projects. All of these projects will be valuable themselves—gamma rays reveal processes that longer-wavelength photons cannot—and will greatly enhance the returns of neutrino and gravitational-wave observatories.

In cosmic rays, the scientific opportunities are also outstanding, including the possibility of eventual directional astronomy with charged particles, but continued science and technology development is needed to drive sufficient advances over current and planned experiments.

For the whole multi-messenger program above, the costs would be modest (details below, in Box L.2 and supporting text) while the scientific returns would be outstanding. Even greater returns will follow if these new observatories are operated simultaneously with each other and with the growing transient program in conventional electromagnetic astronomy, especially the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST). Theory and computing support will be critical to all aspects of this, as those connect disparate observations to each other and to fundamental physics, predict new phenomena, and guide efficient experimental designs to observe them.

Until new discovery-class observatories for gravitational waves, neutrinos, and gamma rays begin operations, it is critical to maintain support for key existing experiments: the Laser Interferometer Gravitational-Wave Observatory (LIGO) and its A+ upgrade, NANOGrav, IceCube, the High-Altitude Water Cherenkov (HAWC) Observatory, the Fermi Gamma-ray Space Telescope (Fermi), the Neil

Gehrels Swift Observatory (Swift), space- and ground-based cosmic-ray observatories, as well as a range of smaller experiments, especially those that are developing new technologies. The combined timeline for current and future facilities is shown later in Figure L.3, and individual capabilities are discussed below.

Developing this new program will require unprecedented coordination of these new projects with each other and with conventional astronomy programs, technology development to create significantly greater capabilities, and cultural work to develop the necessary connections between astronomy and physics and in how collaborations operate.

If this new program is realized, by 2030 the broad field of astronomy would look very different, with robust new astronomies in gravitational waves, neutrinos, and gamma rays built on incredible discoveries individually and especially through their synthesis, allowing detailed studies in the 2030s. As part of this, the United States would maintain and grow leading roles in forefront fields that it had huge roles in developing and which are now rapidly expanding worldwide. Without this new program, the science opportunities—and especially the U.S. roles in them—would be greatly impoverished.

BOX L.1

Classification of Projects by Funding Requirements

As per guidelines for the Astro2020 survey, projects have been classified according to their costs.

For ground-based experiments, small-scale projects cost up to $20 million, medium-scale projects cost $20 to $70 million [e.g., the NSF Mid-Scale Research Infrastructure (MSRI) program], and large-scale projects cost more than $70 million [e.g., the NSF Major Research Equipment and Facilities Construction (MREFC) program].

For space-based experiments, small-scale missions cost up to $0.5 billion (the NASA Astrophysics Pioneers and Explorer programs), medium-scale missions cost $0.5 to $1.5 billion (the NASA Probes program), and large-scale missions cost more than $1.5 billion (the NASA Flagships program). NSF

Taking into account the reports of this Panel and others, the Steering Committee makes recommendations for specific large-scale projects, as well as for general directions that the funding agencies should consider when evaluating medium- and small-scale projects through the normal competitive review process.

L.2 DREAMS OF NEW ASTRONOMIES

The fundamental goal of astronomy is to observe and understand the universe and its constituents. In 2015, the detection of gravitational waves from the collision of two black holes (GW150914) started a new era in gravitational wave astronomy. Two years later, the age of multi-messenger astronomy was ushered in by two breakthrough discoveries: the detection of gravitational waves by LIGO from a binary neutron star inspiral (GW170817) and the detection of an astrophysical neutrino by IceCube during a blazar flare (TXS 0506+056). These source detections, plus IceCube observations since 2013 of a bright, hard-spectrum TeV–PeV diffuse background, have revealed much richer prospects than can be seen with conventional astronomical observations. We now know that our universe contains:

- Extreme gravitators, where the dynamics of strong-field gravity produce deformations of the fabric of space-time that are detectable across cosmic distances. LIGO has observed dozens of electromagnetically dark mergers of binary stellar-mass black holes, showing that gravitational-wave observations are essential to astronomy.

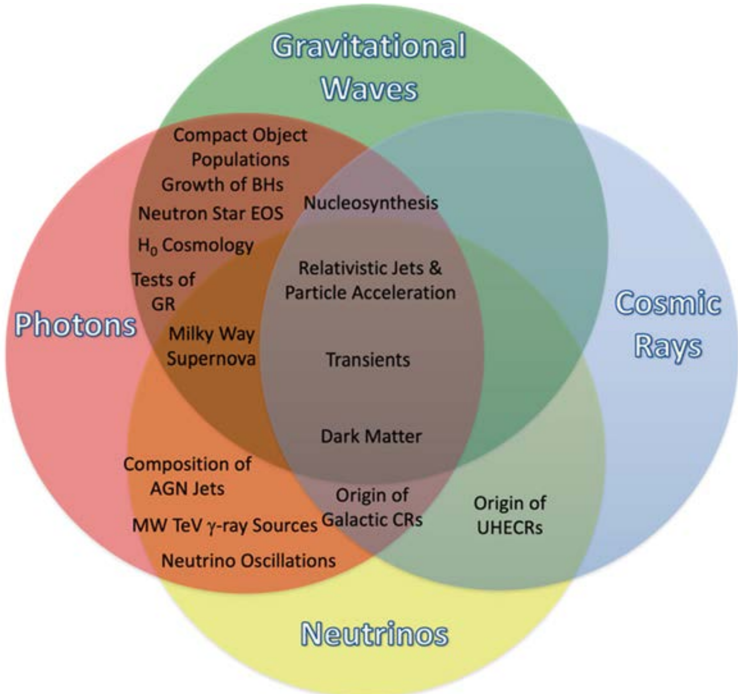


FIGURE L.3 New messengers are essential to solving many longstanding questions, including the high-priority questions identified by the Astro2020 science panels. In this schematic diagram, overlaps indicate where the science is best done with multiple messengers.

- Extreme accelerators, with huge luminosities of charged particles and accompanying gamma rays and neutrinos, and with per-particle energies ranging up to the TeV–PeV range and sometimes much higher. IceCube observations of the diffuse neutrino flux suggest a dominant population of sources that are gamma-ray obscured, showing that neutrino observations are essential to astronomy.
- Many and varied multi-messenger sources, where simultaneous observations are critical. The 2017 observations of GW170817 and TXS 0506+056 achieved their discovery potential thanks to detections by Fermi and other gamma-ray observatories, plus optical and other telescopes, which facilitated quick, deep follow-up in astronomical electromagnetic bands. As the sensitivity of gravitational-wave and neutrino astronomy increases, facilitating the pathways between different types of observatories will become even more essential.

Only observations with new messengers can reveal these new sources and solve many longstanding questions. In astronomy, these questions include the details of stellar endpoints, the jets in active galactic nuclei, and the universe's dark processes. In physics, these include fundamental tests of gravity, the nuclear equation of state, and the particle properties of neutrinos and dark matter. Figure L.1 highlights some of these potential discovery areas.

New-messenger observations are key to answering some of the high-priority questions identified by the Astro2020 science panels. For the Panel on Compact Objects and Energetic Phenomena (COEP), these include their overall discovery area plus several questions:

- *COEPD:* Transforming our view of the universe by combining information from light, particles, and gravitational waves.
- *COEP1:* What are the mass and spin distributions of neutron stars and stellar black holes?

- *COEP2:* What powers the diversity of explosive phenomena across the electromagnetic spectrum?
- *COEP3:* Why do some compact objects eject material in nearly light-speed jets, and what is that material made of?
- *COEP4:* What seeds supermassive black holes and how do they grow?

New-messenger observations are also key to questions identified by the Panel on Cosmology (COS) and the Panel on Stars, the Sun, and Stellar Populations (STARS), especially:

- *COS3:* What physics drives the cosmic expansion and large-scale evolution of the Universe?
- *COS4:* How will measurements of gravitational waves reshape our cosmological view?
- *STARS4:* How do the Sun and other stars create space weather?

To enable discoveries that answer these questions, the first step is to build powerful observatories for new messengers. For sources with multi-messenger signals, such as binary neutron star mergers, flares of active galactic nuclei, and a Milky-Way supernova, broad-based, simultaneous detections are critical. Gamma-ray monitors like Fermi and Swift, but with improved capabilities, are needed to find and localize events for follow-up in conventional electromagnetic observations, which provide precise details and localization in sky coordinates and redshift. Progress depends not only on new experiments, but also on new investments in theory and computation, as well as on addressing broader issues, as discussed in Section L.6.

The increasing tilt of physics research toward astrophysics and cosmology arises from the recognition that these are powerful tools to address fundamental questions beyond the reach of laboratory experiments. Arguably, this tilt has been one of the most significant developments for astronomy in the past few decades, and it can be encouraged. When do observations with new messengers become astronomy *per se*? One answer is when we detect multiple localized sources. For gamma rays and gravitational waves, this has been attained; for high-energy neutrinos, it is within close reach; and for cosmic rays, it is a hope to be nurtured. Sources that appear only in gravitational waves and/or neutrinos are especially interesting, as they reveal the universe's dark processes. By the end of this decade, astounding discoveries are near certain, provided that the field and funding agencies make the right choices. For neutrinos and gravitational waves, observing even small numbers of multi-messenger point sources can be extremely significant, as this could reveal the origin of IceCube's diffuse background and critical details about binary neutron-star mergers and their connection to gamma-ray bursts. Detections in the 2020s can pave the way for much more powerful observatories and higher statistics in the 2030s Figure L.2 illustrates some of the prospects. This report highlights the most critical elements of a bold, broad program that is within reach.

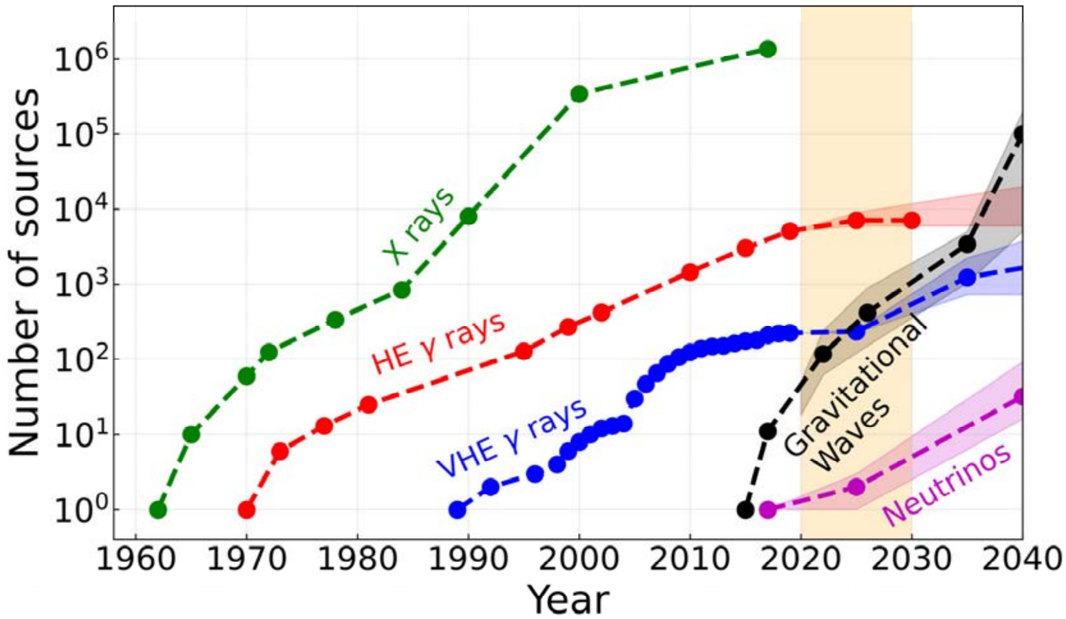


FIGURE L.2 Building on past successes, new astronomies can be opened now. The fields of X-ray and gamma-ray astronomy were built over decades, with the number of observed sources increasing by orders of magnitude. Gravitational-wave and neutrino astronomies are just starting and hold the promise of similar growth. Figure L.4, below, provides more detail.

L.3 DESCRIPTION OF ACTIVITIES CONSIDERED

This panel differs from other program panels in three ways. First, although its topics have long been pursued, its relevance for astronomy has increased greatly in the past several years. Second, many of the proposed investments within its scope would be funded outside the usual mechanisms in astronomy. Third, it is the natural home of multi-messenger astronomy, which requires a new level of coordination among the fields examined by the Astro2020 survey.

The panel inputs included the forefront questions identified by the science panels, 23 project white papers, responses to requests for information sent to many projects, and independent Technical, Risk, and Cost Evaluations (TRACE) (see Appendix O) of some projects. All of these materials were carefully read and considered by the panel over its two in-person meetings and its weekly online meetings. The panel took into account the international context and physics projects that did not submit white papers. It analyzed the capabilities needed to address science-panel questions, through observing and understanding extreme gravitators, extreme accelerators, and multi-messenger sources, in comparison to opportunities. Importantly, the panel considered how to maximize the scientific return by developing diverse fields as a coherent whole on a viable, coordinated timeline.

The fields in the scope of the panel—gravitational waves, neutrinos, gamma rays, and cosmic rays—need to be considered coherently with each other and with the conventional astronomy program to maximize the value of research in each field and for astronomy as a whole. The most compelling programs for new investment thus strongly depend on the landscape of existing and planned experiments, in the United States and abroad, and its gaps. Figure L.3 illustrates the landscape of capabilities. In further detail:

- *Gravitational Waves:* To probe extreme gravitators on a wide range of mass scales, sensitivity in multiple frequency bands is needed. In the audio (kHz) band, LIGO has made several discoveries and has clear upgrade plans in the 2020s but needs technology development for a successor in the 2030s. In the nHz band, NANOGrav has been operating successfully and has set important limits; the continuous long-term timing needed is threatened by decreased funding for the Arecibo[1] and Green Bank Observatories. In the mHz band, LISA will be a powerful new capability in the 2030s; if the United States is to have an important role in this project, it needs to increase its participation in the 2020s. In the future, it will be possible to expand coverage to the 0.1–1 Hz band, which would be important for observing the merger of intermediate-mass black holes with 10^3–$10^4 M_\odot$ masses.

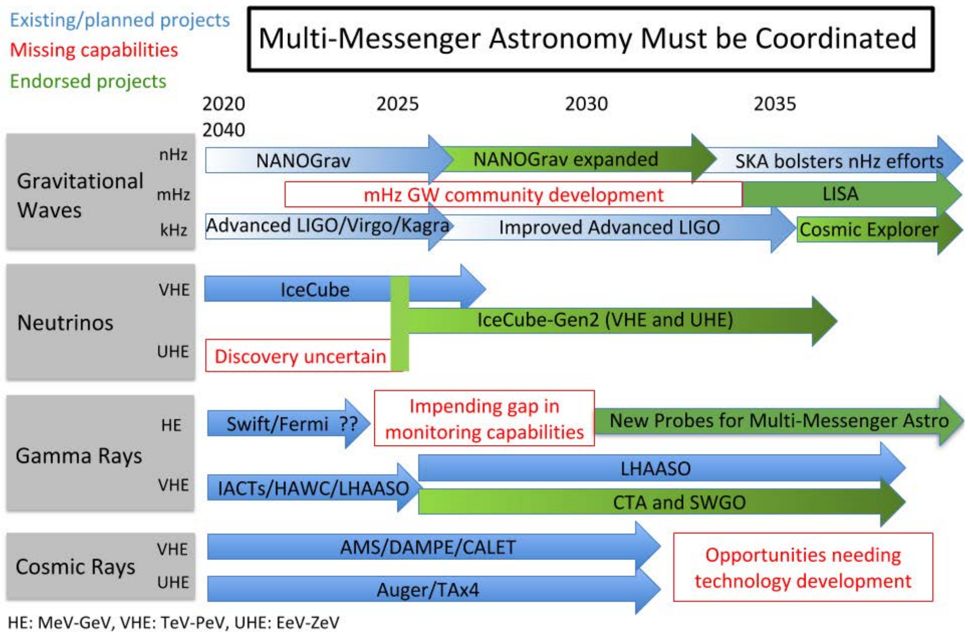


FIGURE L.4 Schematic high-level view of capabilities in different messengers over decades (blue: existing or planned, red: missing capabilities, green: endorsed new projects, dated by construction starts). Gradient shading indicates projects that can start taking data as construction proceeds. Not shown are many promising potential projects for which technology development is needed. With each messenger, the discovery prospects are outstanding; with multi-messenger observations, they could be transformative.

- *Neutrinos:* To probe extreme accelerators, neutrino observatories need greater sensitivity across a range of energies. In the TeV–PeV range (very high energy, VHE), IceCube has detected a bright, hard-spectrum diffuse background and one likely source. In the EeV–ZeV range (ultra high energy, UHE), we know from cosmic-ray data that powerful sources must exist, but experiments have only set limits on the neutrino flux. In both energy ranges, dramatic leaps in sensitivity are needed and are feasible if relevant research is adequately supported. To go further, there are a variety of proposed experiments with promising new ideas for ultra-high-energy neutrino detection.
- *Gamma Rays:* To probe multi-messenger sources, gamma-ray observations are critical both as sky monitors and as precision tools. In the keV–MeV–GeV range (high energy, HE),

[1] The catastrophic loss of the Arecibo facility occurred after the PAG Panel completed its deliberations and presented its findings to the Steering Committee. Following the lead of the Panel on Radio, Millimeter, and Submillimeter Observations from the Ground (RMS), which evaluated such facilities, the PAG Panel did not attempt to revise its report. Please see additional notes in the RMS Panel report.

Fermi and Swift have been indispensable. However, even as increasing investments are made in gravitational-wave and neutrino observatories, these satellites are reaching the end of their lives with no successors planned. In the TeV–PeV range, there are exciting developments worldwide, but without U.S. support for and involvement in these activities, the United States will lose its leadership role.

- *Cosmic Rays:* Although cosmic rays do not point back to their sources, owing to magnetic deflections, observations with the Pierre Auger Observatory, Telescope Array, and other facilities show the existence and high power of extreme accelerators. Further, because cosmic rays are samples of matter from distant sources, their composition information is valuable. Directional astronomy with the ultra-high-energy (UHE) cosmic rays may be possible if a sufficient component of the flux is protons as opposed to nuclei; the flux of ultra-high-energy neutrinos is sensitive to the nature of the cosmic rays, being higher for a lighter composition. Further measurements and technology development are needed to dramatically improve sensitivity for breakthroughs in the 2030s.

These new messengers will be especially powerful when combined with each other and with conventional astronomical observations. A key example is binary neutron star mergers, which will be plentifully observed when the LIGO A+ upgrade achieves its design sensitivity in the mid 2020s. Without sky monitoring capabilities and sensitive gamma-ray detectors in the keV–MeV range, it will be very challenging to localize these events promptly and study them in detail. Without GeV–TeV–PeV gamma-ray and neutrino data, it will be challenging to tell if these short GRBs accelerate cosmic rays. Another example is AGN flares, for which gamma-ray data are critical to determining which sources have high-energy activity. Other examples include a Milky Way supernova, a nearby long GRB, and the tidal disruption of stars by black holes.

L.4 MISSIONS AND PROJECTS ENDORSED FOR THE SURVEY

A successful new-messenger program requires investments in this decade in new observatories for gravitational waves, neutrinos, and gamma rays. The greatest discovery potential comes from having a rich, connected observational program. Multiple discoveries of seminal importance to astronomy are likely. This program is realistic owing to the modest costs of these projects. Success also depends on continued support for present active experiments. In the following, the panel’s view of the most compelling large-, medium-, and small-scale investments are described in turn, selected from the many submitted white papers. Difficult choices had to be made.

BOX L.2 Multi-messenger Program for the 2020s Endorsed by the PAG Panel

- **Large-scale:** IceCube-Gen2 Neutrino Observatory (NSF MREFC program), $345 million.
- **Medium-scale, Gravitational Wave Program:** Investment in three gravitational-wave bands, with support for technology developments toward Cosmic Explorer (NSF), $66 million; increased sensitivity for NANOGrav (NSF), $118 million; and increased U.S. participation in LISA data analysis and science (NASA), $100 million.
- **Medium-scale, Gamma-Ray Program:** Investment in a gamma-ray program both in space, with a new NASA Probe-scale mission, in the range $0.5 to 1.5 billion; and on the ground, with NSF support for participation in the international CTA and SWGO efforts (NSF), $70M and $20M, respectively.
- **Small-scale:** Technology development, plus theory and computation, to enable breakthroughs in future capabilities.

The costs are inclusive but approximate. See text for details.

L.4.1 Large-Scale Investment: Neutrino Program

To develop discovery-class observatories for astrophysical neutrinos over a wide energy range, the panel endorses continued growth in this field under U.S. leadership. The centerpiece would be IceCube-Gen2; compared to IceCube—one of the largest, most successful, and most visible NSF investments—it would have greatly increased sensitivity while having a comparable RY cost. IceCube-Gen2 is designed to resolve IceCube's observed TeV–PeV diffuse background into sources and to open new frontiers at higher energies, up to the EeV–ZeV range. In addition, as discussed in Section L.4.4, the panel also endorses technology development that may facilitate even larger future experiments at those higher energies.

L.4.1.1 NSF: IceCube-Gen2

The IceCube Neutrino Observatory was constructed at the South Pole during 2004–2010; data-taking and analysis with newly deployed hardware followed soon after each year's commissioning. The capabilities of the original IceCube facility are being improved upon by the IceCube Upgrade, a relatively small project that will be completed in 2023 that adds a dense infill of optical sensors to improve sensitivity at low energies and to better calibrate IceCube. IceCube-Gen2 will create a third-generation observatory for high-energy neutrinos sited at the South Pole, greatly improving upon the capabilities of IceCube and its predecessor, the Antarctic Muon and Neutrino Detector Array (AMANDA). Construction could start in 2024 and would take about 10 years to complete, during which the experiment would be taking data with an increasingly larger detector. The primary component of IceCube-Gen2, intended for TeV–PeV neutrinos, is an array of optical sensors, much like IceCube but with significantly enhanced sensitivity and directionality. The secondary component ($\simeq 10$ percent in cost), intended for higher-energy neutrinos, is an array of radio sensors, building on the heritage of several experiments, but with dramatically increased sensitivity. At present, there are no funded detectors worldwide that could compete with IceCube-Gen2, which is led by the United States, in either energy range.

Scientific Context—More than a century since the discovery of charged cosmic rays, their origins are still unknown, owing to magnetic deflections that obfuscate their sources. The dominant cosmic-ray component is nuclei, principally protons. From observations of high-energy gamma-ray sources, there are

many proposed sites for where cosmic rays (including electrons, which are subdominant) are accelerated. However, there is an essential question that has not been resolved. Where nuclei are accelerated, both gamma rays and neutrinos are produced through hadronic processes, especially pion production and decay. Where electrons are accelerated, only gamma rays are produced, through leptonic processes, especially inverse-Compton scattering. Detecting neutrinos from some source class would be definitive evidence of efficient hadronic acceleration. Together with sufficiently sensitive nondetections of neutrinos from other source classes, which would favor efficient acceleration of electrons only, this would revolutionize our understanding of cosmic-ray origins and the nature of gamma-ray sources, each a long-standing mystery.

Starting in 2013, IceCube has detected a diffuse, hard-spectrum, TeV–PeV flux of astrophysical neutrinos, now with $\simeq 60$ events of energies $\gtrsim 60$ TeV that are distinct from the steeply falling foreground of atmospheric neutrinos that dominates at lower energies. The diffuse background is extragalactic, with a Milky Way component surprisingly absent. Comparison to the diffuse gamma-ray background suggests that the neutrino sources may be dominantly gamma-ray dark, which would make the results even more exciting. In 2017, IceCube detected one likely source, the blazar TXS 0506+056, with one high-energy neutrino detected in association with a gamma-ray flare (GeV by Fermi, TeV by MAGIC), with a significance of 3σ. IceCube then also detected, at a significance of 3.5σ, a 2014–2015 neutrino flare without accompanying gamma-ray emission, from the same source.

IceCube-Gen2 is expected to make several discoveries that go beyond the capabilities of the existing IceCube (including the Upgrade). First, IceCube-Gen2 would dramatically improve on the existing spectrum, skymap, and flavor information of the diffuse data, improving statistics by an order of magnitude, leading to important clues about the origin of these neutrinos. Second, IceCube-Gen2 has the power to resolve the diffuse extragalactic background into discrete sources, for which the case is based on the luminosity and number density of sources, and does not depend on the TXS 0506+056 source, and to make first detections of Milky Way sources, for which the case is based on high-energy gamma-ray observations. Third, by improving present sensitivity at higher energies by about two orders of magnitude, IceCube-Gen2 would enable first detections at such energies, including of the cosmogenic flux.

The cosmogenic neutrino flux, reaching the EeV–ZeV range, is owing to interactions of ultra-high-energy protons with the cosmic microwave and infrared backgrounds. There is strong circumstantial evidence for these interactions, with a 20 σ pileup feature below $\sim 10^{19.5}$ eV in the cosmic-ray spectra observed by the Pierre Auger Observatory and the Telescope Array. However, this feature could also indicate the breakup of nuclei. The neutrino flux is sensitive to the cosmic-ray composition, being larger for protons compared to nuclei, encouraging the development of new neutrino observatories to test the cosmic-ray composition. A light composition could open a window for future directional astronomy with charged cosmic rays. Encouraging the development of new neutrino observatories, with much larger statistics than IceCube-Gen2, to probe the most extreme accelerators and the cosmic-ray composition through measuring the neutrino spectrum over a broad energy range.

IceCube-Gen2 will provide critical input to the Astro2020 science questions COEPD, plus some of COEP1, COEP2, COEP3, COEP4, and STARS4, depending on what the sources of the observed neutrinos are. More generally, the first definitive source detections in high-energy neutrinos would have a huge, broad impact, much like that for the first detection of a binary neutron star merger in gravitational waves. Contemporaneous observations of IceCube-Gen2 and other facilities—including those for gamma rays (CTA, based in Europe, and the Large High Altitude Air Shower Observatory [LHAASO], based in China), gravitational waves (worldwide, but especially LIGO and its upgrades), and transient astronomy (worldwide, but especially the LSST project)—would have synergistic benefits that could lead to breakthroughs.

Implementation—The optical array of IceCube-Gen2, intended for the detection of TeV–PeV neutrinos, is principally a larger version of IceCube, for which the technology and science have been demonstrated. The instrumented volume would be 7.9 km^3, with 120 vertical strings of 100 optical sensors buried in deep (2.6 km) holes with a horizontal spacing of 240 m. For IceCube, the comparable

figures are 1 km^3, 86 strings each with 60 sensors, and with a typical spacing of 125 m. IceCube-Gen2 would use improved optical sensors: instead of a single 8” photomultiplier tube per module, there would be 24 grouped 3” tubes. These differences provide benefits in flux sensitivity, angular resolution, and cost. The panel’s assessment is that the proposed size of the IceCube-Gen2 optical array is the minimum needed to ensure detection of point sources, the principal goal.

The radio array of IceCube-Gen2, intended for the detection of higher-energy neutrinos, builds on decades of technology development to make a huge leap in sensitivity. The proposed array design comprises ~200 stations deployed over an area of 500 km^2. Each station would be instrumented with both horizontally and vertically polarized receiver antennas, deployed on the surface and/or to a depth of 100 m. This would build on the heritage of previous experiments, including RICE, ARA, ARIANNA, and ANITA.[2] There is a compelling case to include a radio array as part of IceCube-Gen2, as discussed below. The panel’s assessment is that the proposed size of the IceCube-Gen2 radio array is the minimum needed to make a first detection of cosmogenic neutrinos in nominal scenarios, the principal goal.

It is critical to have both the optical and the radio arrays to cover a wide range of energies and to cross-calibrate. Different spectral components are expected in the TeV–PeV range and at higher energies, and how they connect will be a powerful test of source properties and cosmic-ray composition. The combined detector and its calibrations will also allow unprecedented measurements of ice properties, at wavelengths from 100 nm to 1 m, which will have broader impacts for glaciology.

Delaying the proposed timeline would endanger achieving the scientific goals. Much of the project, installation, technical and scientific expertise needed for IceCube-Gen2 dates back to IceCube and is currently being used and expanded for the IceCube Upgrade; attrition of this personnel base would require the development of a new cadre of specialists. Particularly irreplaceable is the drilling team, which operated the 5 MegaWatt Enhanced Hot Water Drill for IceCube. This expertise has been sustained by the Deep Core and Upgrade efforts; a large gap after the end of the Upgrade would likely result in the loss of this team and a significant delay in commissioning the hardware. A limiting factor in how fast the detector can be built is the short duration (~10 weeks) of the drilling season, which makes it critical to have an experienced, efficient team to complete the detector within the projected budget.

Costs/Risks—The project-estimated cost of IceCube-Gen2 is $345 million in fiscal year (FY) 2020 dollars ($420 million in real-year [RY] dollars). This is approximately the same RY cost as IceCube. As with IceCube, funding for construction of IceCube-Gen2 would be sought through the NSF MREFC program, with support for operations through NSF Division of Physics and the Office of Polar Programs, plus substantial international contributions (much of which is already committed). The MREFC project investment would be stretched out over 10 years owing to the short South Pole construction season; the peak annual MREFC funding required would be ~$50 million in 2024.

IceCube-Gen2 is designed around mature technology. Since the initiation of neutrino telescopes at the South Pole (AMANDA construction began in 1993–1994 after 3 years of prototyping), the techniques for drilling and deploying a distributed optical array to 2500 m depths have been honed to the point that > 99 percent of the optical sensors initially deployed for IceCube show no loss in performance after a decade of data-taking. Radio-based neutrino detection has developed similarly, with the first deployment of hardware at the South Pole in 1995. In addition to their dedicated physics programs and goals, the IceCube Upgrade at the South Pole and the Radio Neutrino Observatory in Greenland (RNO-G) will prototype a significant fraction of the hardware planned for IceCube-Gen2.

The costs of IceCube-Gen2 are significantly reduced by the ability to leverage the infrastructure built by IceCube, including the deployment and operations experience, the data-processing and analysis tools, and the established collaboration. In addition, there are important technological developments, plus the field experience to know that a larger instrument spacing is feasible for the observed hard-spectrum flux. The leverage to reduce costs is especially true for the radio array (about 10 percent of the total costs of IceCube-Gen2), which would benefit greatly from the above plus by sharing facilities for electronics

[2] RICE, Radio Ice Cherenkov Experiment; ARA, Askaryan Radio Array; ARIANNA, Antarctic Ross Ice-Shelf Antenna Neutrino Array; ANITA, Antarctic Impulsive Transient Antenna.

and personnel at the South Pole. Although the radio technology is newer than the optical technology, it builds on substantial heritage and its development would be accelerated by having a larger collaboration. Achieving comparable sensitivity in a dedicated radio array at another site and/or with a different collaboration would be substantially more expensive (but see the discussion of continued technology development in Section L.4.4).

An independent, external Technical, Risk and Cost Evaluation (TRACE) review was done to assess whether IceCube-Gen2, accounting for all risks, had been properly costed to achieve the desired scientific goals on the stated timeline. The TRACE review found that the programmatic risk (science and costs) and schedule risk of IceCube-Gen2 are both medium-to-low. The TRACE cost estimate is 20 percent higher than the project-estimated cost, which the panel considers only a minor concern. The MREFC review process will lead to more accurate accounting. The MREFC-funded IceCube project was constructed with a fixed budget; the project was able to save costs and deploy an increased number of strings, for which there was a powerful science motivation. It is likely that IceCube-Gen2 could do the same. The TRACE projected a schedule that is 7 months longer than that estimated by the project, which again the panel considers only a minor concern.

L.4.2 Medium-Scale Investments: Gravitational-Wave Program

To develop discovery-class, multi-band experiments in gravitational waves, the panel endorses both the continued growth in sensitivity of current gravitational-wave observatories and the development of new ones, in multiple gravitational-wave bands. On the ground, this includes planned upgrades to the LIGO facilities and technology development for its successor, Cosmic Explorer, and continuity and growth of NANOGrav observations. In space, this includes an increased U.S. presence in the science of the LISA mission. The United States has played a key role in the conception of all of these efforts and is currently either leading them or contributing critical input. LIGO is another one of NSF’s largest, most successful, and most visible investments. NANOGrav, whose data set is currently dominated by the Arecibo and Green Bank observatories, provides the bulk of the sensitivity to the International Pulsar Timing Array (IPTA). LISA was initially conceived as a partnership between NASA and ESA; ESA now leads the project, but key technology and analysis are under development in the United States with NASA support.

L.4.2.1 NSF: Cosmic Explorer

Cosmic Explorer is the U.S. component of a future network of third-generation, ground-based gravitational-wave detectors. In the current plan, Cosmic Explorer will be built on the same principles as LIGO, but with 10 times longer arms (40 km) and additional technological upgrades that will provide the ability to measure and characterize every stellar-mass black hole merger in the universe. The corresponding European-based project is the Einstein Telescope, with different design and implementation but with comparable sensitivity and time scale. Cosmic Explorer and Einstein Telescope will be part of a detector network to provide source localization and coverage—critical ingredients for multi-messenger science. The details of how such a network will operate are still to be defined.

Scientific Context—At its current sensitivity, LIGO is able to detect gravitational-wave signals on a roughly weekly cadence. Once LIGO detectors achieve the sensitivity planned for their A+ upgrade, which is expected in this decade, their detection rate will increase by a factor of about 10. The third generation of observatories is intended to increase this rate by a factor of 1000 or more. By lowering the low-frequency sensitivity limit of LIGO from 10 to 5 Hz and reducing the noise by a factor of 10, Cosmic Explorer will reach gravitational-wave signals all the way back in cosmic time, for a powerful and diversified science program. The science goals for Cosmic Explorer include determining the nature of the densest matter in the universe, enabling multi-messenger observations of binary neutron star systems, and

measuring the geometry and expansion rate of the universe independent of electromagnetic observations. Cosmic Explorer will also provide insights into the evolution of massive stars, the physics of supernovae, and the origin of pulsar glitches, and maybe find exotic sources. Cosmic Explorer will provide critical input to the Astro2020 science questions COEPD, COEP1, COEP2, COEP3, COEP4, COS3, and COS4.

Implementation—Cosmic Explorer will be an L-shaped laser interferometer built on the surface of geologically appropriate and seismically quiet land in a U.S. location still to be identified. Cosmic Explorer Stage 1 (CE1) will adopt technology tested in the 4 km LIGO facilities to a 40 km detector; the longer arms will increase the amplitude of the observed signal but not the noise, thus providing better sensitivity. Cosmic Explorer Stage 2 (CE2) will provide further sensitivity improvements with new technology that will mitigate quantum and thermal noise. This plan spans multiple decades and will ultimately be a large-scale program, but the requested investment in this decade is a medium-scale ground-based investment from NSF to implement technology developments that will proceed in parallel with the approved LIGO A+ upgrade. These developments will be needed to initiate observations with CE1 in the 2030s and CE2 in the 2040s. The strategy of performing technology development for future detectors while maintaining the operations of current detectors worked very well for the transition from the initial proof-of-concept LIGO to the currently operational LIGO; the same model is expected to work for the transition from the current facilities to Cosmic Explorer, with the bonus that technology being planned for the future may end up being affordable enough to be first installed in the current 4 km facilities.

Costs/Risks—The Cosmic Explorer request in this Decadal Survey is for medium-scale funding for a design study, at the level of $65.7 million in FYs 2020 to 2025: $33 million for two engineering studies, $20 million for a prototype CE chamber, and the remainder for other upgrades and for governance. Based on 2011 estimates for the Einstein Telescope and the historical cost of LIGO, the panel expects this project will become a large-scale investment in the 2030s. The requested investments in the 2020s will yield results from engineering studies and experiments that will then help produce solid estimates for the costs and risks of the full Cosmic Explorer project. In addition, Cosmic Explorer is a ten-fold expansion of an experiment that has been demonstrated to work, and it is proposed by a team with world-class expertise and leadership in its field.

L.4.2.2 NSF: North American Nanohertz Observatory of Gravitational Waves (NANOGrav)

NANOGrav regularly observes 75–200 millisecond pulsars to detect and characterize gravitational-wave emission in the nHz band via the “pulsar timing” technique. Correlated changes in the arrival times of the pulses are analogous to correlated changes in laser phase in LIGO or LISA; NANOGrav is therefore a galactic-scale gravitational-wave interferometer with arm lengths of hundreds of parsecs. NANOGrav is currently funded as an NSF Physics Frontier Center. The IPTA is a consortium of pulsar timing arrays (PTAs), including NANOGrav, that shares data to increase the sensitivity to nHz gravitational waves. NANOGrav, and specifically the Green Bank and Arecibo[3] Observatories, provide the bulk of the sensitivity to the IPTA. NANOGrav has a strong tradition of conducting training programs for young scientists and significant outreach programs.

Scientific Context—The nHz gravitational-wave band is the only way to measure the cosmic merger rate of 10^8–$10^9 M_\odot$ black-hole binaries, a critical test of the evolution of both black holes and galaxies. NANOGrav has placed limits on the nHz stochastic gravitational-wave background, already constraining current models of supermassive black-hole binary formation. Under current projections, NANOGrav expects to detect the stochastic background owing to supermassive black-hole binaries in the first half of this decade and to detect individual sources in the second half of this decade. The amplitude of the background will measure the scaling of black-hole mass with host-galaxy mass as well as dynamical-friction time scales. Further measurements—the spectrum of this background, in combination

[3] See note 1, above.

with the detection of individual 10^8–$10^9 M_\odot$ binary mergers— ill yield a definitive picture of how these supermassive black holes evolve in their galactic environments. With this, NANOGrav will provide crucial input on the Astro2020 science questions COEPD, COEP4, and COS4.

Implementation—To fulfill on not just first detections, but also to make detailed measurements of the nHz gravitational-wave band (both stochastic background and single sources), NANOGrav requires continued access to the Green Bank and Arecibo observatories, in addition to an expansion of their capabilities with future radio facilities with larger collecting area. This expansion is crucial to achieving the daily cadence of observations that NANOGrav requires to detect and characterize single continuous-wave sources of gravitational waves. With the Arecibo and Green Bank observatories alone, NANOGrav has sufficient collecting area but insufficient overall observing time for detecting single sources.

Costs/Risks—The budget for NANOGrav science is $118 million over the decade of the proposed plan. While the cost exceeds the nominal ground-based medium-scale range (Box L.1), the panel considered it in this range based on its present scale and the possibilities of either changes in scope or attracting investments from other partners. These costs are dominated by telescope usage and personnel. The observational capabilities of NANOGrav would be greatly increased by the realization of new and expanded radio facilities; an analysis of the feasibility of such facilities was not in the scope of the panel but is included in the report of the Astro2020 Panel on Radio, Millimeter and Submillimeter Observations from the Ground (RMS). In the absence or delay of such facilities, the defunding of Green Bank and Arecibo would place at risk NANOGrav science. The sensitivity increases as the time baseline of observations increases, so the loss of a telescope is not just a loss of sensitivity at the time of its loss, but also the deprecation of the data set for its long-wavelength capabilities. The Green Bank Observatory is operated as a partnership between federal, state, and private sources. The $118 million NANOGrav budget for the decade includes $30 million for telescope time on these two telescopes, and allows NANOGrav to use 750 hours per year, or about 10 percent of the on-sky time, on each telescope. That time is still a fraction of the total telescope time required by the project. In the ideal scenario specified in the NANOGrav white paper, the required time would come from potential future facilities (about one half), Green Bank and Arecibo (about one-third), and the remainder (about one-sixth) from a combination of IPTA telescopes.

L.4.2.3 NASA: Laser Interferometer Space Antenna (LISA)

LISA is a project led by the European Space Agency (ESA), with significant contributions anticipated from several ESA member states and NASA. LISA will be the first space-based gravitational-wave detector, sensitive to the mHz range. The 2020s will be a crucial decade for LISA, as activities ramp-up in preparation for a launch in 2034.

Scientific Context—Multi-messenger astronomy cannot reach maturity without an instrument capable of observing gravitational waves in the mHz band, which are emitted by some of the most interesting sources. Because of seismic noise, which dominates below 1 Hz, ground-based gravitational-wave detectors cannot be sensitive to mHz signals; however, LISA will have this capability by virtue of being space-based. LISA will be able to observe all merging supermassive black holes in the universe (10^5–$10^7 M_\odot$ masses), the inspiral of small compact objects into supermassive black holes to redshifts of order one, white dwarf and neutron star binaries in the Milky Way, and stochastic backgrounds from the early universe. LISA will also study the dynamics of dense nuclear star clusters and explore the fundamental nature of gravity and black holes, as well as shed light on the existence of ultralight boson fields, a dark-matter candidate that would grow around black holes through superradiance. Last but not least, LISA will probe the cosmic expansion rate, use stochastic gravitational-wave backgrounds to understand the early universe and TeV-scale particle physics, and listen for gravitational-wave bursts from serendipitous sources. LISA will provide critical input on the Astro2020 science questions COEPD, COEP2, COEP3, COEP4, COS1, COS2, COS3, and COS4.

Implementation—The panel was not asked to directly evaluate the LISA mission, as this is an already-approved ESA mission with NASA partnership, with a planned launch in 2034. Instead, the panel was asked to evaluate the possibility of changing the scope of LISA funding. The panel endorses investments in increasing the scope of U.S. participation in LISA, with two categories of contribution:

- (Contribution 1) An increase in U.S. LISA science funding, with a decade-long, dedicated NASA program that researchers at universities and other centers could reliably count on to develop LISA-specific gravitational-wave data analysis tools, gravitational-waveform models, and science-extraction techniques.
- (Contribution 2) Support for a U.S. LISA Science Facility to (a) implement and coordinate the U.S. role in the ESA-led project-level data analysis, (b) provide outside users at universities and research institutes access to mission data at a variety of levels, and (c) provide tools to facilitate working with LISA data and for combining LISA data with other facilities in multi-messenger investigations.

These two items are complementary, as the science facility would integrate the research sponsored by the dedicated U.S. funding program in the ESA-led program, which only supports science development within Europe. This increase in the scope of U.S. efforts would not duplicate other ESA efforts, but rather would add a critical contribution to the development of tools and of the analysis framework that will be needed to extract the most science from the data by 2034. It would also empower U.S. scientists to work with LISA data, as they do with other observatory facilities, and overall strengthen the role and input of U.S. scientists in the LISA mission. A lack of support for U.S. scientists to work on LISA science (modeling, data, and computation) would severely diminish the U.S. capabilities in the future of gravitational-wave science.

The analysis of LISA data will be unlike that of LIGO and NANOGrav: the three instruments are based on distinct technologies, observe the gravitational-wave spectrum in nonoverlapping frequency ranges, and have unrelated noise sources, therefore observing different sources with different challenges. The unique character of LISA data requires new gravitational-wave models (both analytical and computational) and new analysis techniques. Just as an example, space-based detectors will routinely observe gravitational waves emitted by binary systems with intermediate and sometimes extreme mass ratios that inspiral in generic (eccentric and double spin-precessing) orbits, and the numerical, analytical or semi-analytical models currently in existence either cannot describe such systems at all, or are, at best, not sufficiently accurate yet. Different groups of scientists in the United States are currently (and have been for more than 30 years) been working on these nonoverlapping techniques.

Costs/Risks—NASA is currently supporting a range of potential contributions to LISA including instruments, spacecraft elements, and science analysis, in the medium-scale range of $400 million to $600 million. This would cover the cost of the hardware deliverables, as well as contributions to the science ground activities, the U.S. Guest Investigator programs, and NASA project overhead, including management, systems engineering, project science, and mission assurance. The suggested increase in LISA support from NASA would be $100 million for the decade, with $30 million to $40 million for sponsored science funding and $50 million to $60 million for the U.S. LISA Science Facility.

The scope of NASA's LISA Preparatory Science program encompasses the science program described above, but at a lower level of effort, and it is not clear if the Preparatory Science program will be continued throughout the decade. Regarding the increase in U.S. LISA Science funding (Contribution 1), dedicated, reliable, and decade-long funding for a LISA science program at the level of about $3 million to $4 million per year ($30 million to $40 million per decade) would be in line with the dedicated funding channel established by the NSF Gravity program for LIGO research support, which was instrumental for LIGO's success; a similar need can be envisioned for LISA. This new dedicated, decade-long funding program would enable support for roughly seven research groups on 3-year, $500,000 per year grants that researchers could count on over the decade to expand the field of gravitational-wave astronomy and the U.S. role in it.

Decade-long funding of $50 million to $60 million for a U.S. LISA Science Facility (Contribution 2) would enable U.S. community participation in a new and unexplored regime that requires new tools, techniques, and simulations, and is comparable to previous support for other science center activities in the United States. Based on past experience of NASA deploying missions using new messengers, significant investments have been needed to adequately prepare the astronomical community to take full advantage, even in cases where there is substantial heritage from prior missions.

L.4.3 Medium-Scale Investments: Gamma-Ray Program

To develop discovery-class capabilities in multi-messenger astronomy, the panel endorses a dedicated Probe-scale space mission with a suite of capabilities for these sources, plus U.S. participation in international ground-based gamma-ray observatories. The timely deployment of a space-based high-energy gamma-ray mission is critical to take advantage of anticipated discoveries from LIGO and IceCube and their successors. The VHE gamma-ray band, which probes the most energetic particles in the universe, can only be accessed from the ground. The panel endorses U.S. participation in the Cherenkov Telescope Array (CTA) and the Southern Wide-Field Gamma-Ray Observatory (SWGO) as VHE observatories that extend source sensitivities to fainter sources, higher redshifts, and faster emission time scales, providing complementary catalogs of sources that span distance scales from the Milky Way to the cosmos.

L.4.3.1 NASA: Probe-Scale Mission for Multi-Messenger Sources

Gamma-ray observations play a critical role in understanding extreme gravitators and extreme accelerators. A Probe-scale mission dedicated to the study of multi-messenger sources would provide wide-field multiwavelength observations, at keV-MeV-GeV energies, at the sensitivities needed to achieve multi-messenger discoveries and to directly answer questions about compact objects and stellar astrophysics. The ideal mission would provide rapid alerts and sky localization for transient sources, enabling timely follow-up observations by other telescopes with narrower fields of view.

Scientific Context—Space-based gamma-ray observations are needed to probe astrophysical sources at high energies, where nonthermal activity is easily distinguished and indicates extreme physical conditions and possibly cosmic-ray acceleration. Continuum gamma-ray observations provide unique information on the structure and composition of relativistic winds and jets in sources such as pulsar wind nebulae, active galactic nuclei, supernovae, and gamma-ray bursts. To match the growing sensitivity of gravitational-wave observatories like LIGO and neutrino facilities like IceCube, advances in gamma-ray sensitivity are urgently needed to increase detection rates and distance horizons. Increasing the populations of well-studied cosmic accelerators and multi-messenger events will require the wide-field, high-cadence, all-sky monitoring made possible by space-based missions.

A space-based mission is also needed to place multi-messenger observations into the broader astronomical context. The combination of gravitational-wave and gamma-ray observations for GRB 170817A bracketed the inspiral of the binary neutron stars and the first emergence of light from the resulting burst, enabling constraints on theories of gravity and initiating a massive campaign of groundbreaking follow-up observations that probed aspects such as heavy-element formation. The association of a neutrino, X rays, and gamma rays from the TXS 0506+056 blazar demonstrated the combined use of these observations to peer into the workings of relativistic jets. Sky localization and rapid-alert capabilities enable the detection of electromagnetic emission by telescopes with narrower fields of view (e.g., radio, optical, X-ray, or VHE gamma rays). The identification of host galaxies (e.g., as possible at radio or optical wavelengths) is crucial to distance determination, standard-siren cosmology to measure the expansion rate, and population studies to constrain source-formation channels. A Probe-

scale mission for multi-messenger sources will provide critical input to the Astro2020 science questions COEPD, COEP1, COEP2, COEP3, COEP4, COS3, COS4, and STARS4.

Implementation—A Probe-scale mission opportunity in the next decade focused on multi-messenger astronomy would provide key capabilities for multi-messenger discovery. Competed instrumentation on a multi-messenger-themed Probe would allow evaluation among options for capabilities. Launching this in this decade is critical for this science area owing to the necessity of coordination with other planned programs (see Section L.3). A single Probe mission cannot meet all the needs for an ambitious program of discovery in multi-messenger astronomy. In addition to a Probe dedicated to this topic, it is important to enable supporting capabilities for multi-messenger astronomy where possible in the implementation of other NASA astrophysics missions, including Flagship missions.

Costs/Risks—The panel reviewed submissions presenting gamma-ray Probe-scale mission concepts capable of providing key multi-messenger capabilities (e.g., AMEGO, APT, and TAP) primarily for their scientific importance to particle astrophysics and gravitation. The existence of several concepts with a high level of technical readiness and mission plans designed to meet the Probe schedule and budget demonstrate the feasibility of addressing this need with a Probe-scale mission.

L.4.3.2 NSF: U.S. Participation in the Cherenkov Telescope Array (CTA) and the Southern Wide-Field Gamma-Ray Observatory (SWGO)

Astronomical observations in the VHE gamma-ray energy band probe the sites of cosmic-ray acceleration. These observations require a combination of high-resolution and high-sensitivity measurements on point sources as well as full-sky monitoring capabilities. A combined program of imaging atmospheric Cherenkov telescopes (IACTs) and particle detector observatories will create a powerful synergistic capability. The nexus of these facilities is critical to the understanding of the emission of the highest-energy nonthermal radiation extending from galactic to our nearby cosmological neighborhood, and to support simultaneous multi-messenger observations with upgraded gravitational-wave and high-energy neutrino astronomy capabilities.

Scientific Context—Ground-based observations of VHE gamma-ray sources provide a critical probe of nonthermal processes in extreme astrophysical environments, such as gamma-ray bursts, pulsar wind nebulae/supernova remnants, jet emission in active galactic nuclei, and accretion disks near Be stars/binary systems. The CTA IACT observatory will provide seasonal pointed observations over modest fields of view distributed across the full astronomical sky. CTA's large detection area (up to 1km^2 in each hemisphere) provides the large photon statistics necessary for observing VHE spectra from extragalactic sources with weak emission (e.g., the starburst galaxy M82), exploring fast emission time scales (VHE flares from gamma-ray bursts like 1901140C, which lasted for tens of seconds), and detecting morphological variations in extended objects (e.g., the supernova remnant IC443). CTA's low energy threshold extends the cosmological horizon for VHE astronomy beyond a redshift of one, thereby enabling VHE observations into the peak epoch of activity for active galactic nuclei and gamma-ray bursts.

High-altitude arrays of particle detectors—such as HAWC, LHAASO (2021 completion), and SWGO—allow wide field-of-view, continuous all-sky monitoring of the visible sky. SWGO and LHAASO can provide an order of magnitude increase in gamma-ray sensitivity for similar observations, critical for triggering on transients outside the narrow fields of view of IACTs. Particle detector observatories also have superior ability to detect extended, diffuse sources such as supernova remnants and pulsar wind nebulae (e.g., Geminga), the Galactic Plane, and complex, extended-emission regions containing dense clusters of astrophysical sources and phenomena (e.g., the Cygnus region). Using this capability, HAWC discovered a population of previously unknown, angularly extended, hard-spectrum Milky Way sources in the multi-TeV energy range around pulsar-wind nebulae (called "TeV halos").

The CTA observatory, consisting of both Northern and Southern Hemisphere sites, provides access to the highest-sensitivity pointed observations across the full astronomical sky, with the highest

energy and angular resolution. SWGO's Southern Hemisphere sky coverage will complement the Northern Hemisphere coverage of the LHAASO observatory, providing daily coverage of the full sky at very high energies. The combination of CTA and LHAASO/SWGO provides an integrated observational capability that maximizes the scientific opportunities for all-sky multi-messenger astronomy. The success of the broad U.S. program in multi-messenger astrophysics would be greatly enhanced by access to these world-leading facilities. The development of these facilities depends critically on decades of U.S. investment that cannot be capitalized upon without continued U.S. involvement.

Implementation—The international CTA observatory has been under development for more than a decade. U.S. participation in CTA was recommended as a ranked priority in the Astro 2010 Decadal Survey. Since then, the U.S. CTA group has developed and built a prototype medium size (9.7 m diameter) two-mirror Schwartzchild-Couder IACT Telescope (SCT) using funds from the NSF Major Research Instrumentation (MRI) program. The prototype SCT detected the Crab Nebula in Spring 2020 with a partially filled focal plane. The prototype SCT focal plane will be fully populated by 2022 through a second NSF MRI grant. The United States will contribute 10 SCT telescopes to the larger CTA array, which will roughly double the number of medium-scale telescopes. Detailed studies show that this U.S. contribution would dramatically enhance many of CTA's Key Science Projects, ranging from studies of astrophysical sources to searches for dark matter annihilation signals.

The HAWC observatory has demonstrated the synergistic capabilities that high-altitude particle detectors provide to IACT arrays. HAWC also developed the use of a large, modular water-Cherenkov detector design that is scalable to larger arrays and higher altitudes. SWGO is based on this modular design. The increased size of SWGO requires a large international collaboration to manage the construction, operation, and data analysis.

Costs/Risks—The design of the CTA observatory is mature, including a detailed science case, completed site acquisition, optimized observatory design, and prototype testing of every telescope in the array. The project is refining a multi-level work breakdown structure for cost and schedule, a project execution plan, and a plan for assessing and mitigating project risks to cost and schedule. The cost of CTA is estimated to be $500 million (with a U.S. contribution of $40 million) for construction and $3 million per year for U.S. operations, which is well below the $100 million U.S. contribution envisioned in the Astro2010 decadal survey.

The design and costs of SWGO are at the initial stages, with an SWGO construction cost estimate of $60 million (with a U.S. contribution of $20 million) based on extrapolations of cost and schedule of the HAWC observatory. The international SWGO collaboration has only recently been formed (2019) and is in the process of selecting the observatory site and developing a more detailed instrument cost and schedule.

The technologies for both projects are well understood and demonstrated, and so the risks to cost and schedule of both projects are modest. Significant delays in construction funding will result in missed opportunities for U.S. participation in multi-messenger science. Because these projects are being led by international partners, the costs of U.S. participation are greatly reduced.

L.4.4 Small-Scale Investments: Technology Development for Improved Capabilities

In addition to the large- and medium-scale investments described above, it will be critical to support new small efforts that foster new ideas and have the potential to become future larger implementations.

L.4.4.1 Future Observatories for Neutrinos and Cosmic Rays

Highly sensitive neutrino observatories will be needed at ultra-high energies to probe of the origins and composition of ultra-high-energy cosmic rays, building on a possible first detection of the

cosmogenic neutrino flux by IceCube-Gen2. These will allow precise measurement of the spectrum and other properties of the cosmogenic flux—probing the cosmic evolution of the accelerators of ultra-high-energy cosmic rays—if the flux is as large as hoped for, or, at a minimum, an overall flux measurement if the flux is smaller than expected. In detection, these neutrino events will probe center-of-momentum frame energies well above that of the Large Hadron Collider (e.g., for $E_\nu \sim 10^{19}$ eV, $\sqrt{s} \sim 100$ TeV). To develop detectors with exposures well beyond that of IceCube-Gen2 requires funding technology development now. Most of these projects have very modest costs and potentially huge payoffs. As these projects are led by physics groups, not all submitted white papers to the Astro2020 survey.

At present, all proposals for detecting cosmogenic neutrinos are based on radio instrumentation of natural formations, as radio has tremendous advantages of long attenuation lengths and sensitive, low-cost detectors. The proposed projects involve sites worldwide, at least for the development phase. Ultimately, some may propose to locate in Antarctica or even the South Pole specifically, but not necessarily. Neutrino interactions lead to energetic particle showers. In a dense medium like ice, a shower emits coherent radio signals through the Askaryan effect. In a tenuous medium like air, neutrino interactions in a sufficiently thick nearby mass such as a mountain can lead to tau leptons that decay in air, producing extensive air showers that can be detected via radio signals emitted by geomagnetic synchrotron processes. In addition, active detection of showers in dense media through radar is possible owing to the reflectivity of radio waves on the ionization the showers leave behind, recently measured for the first time in a laboratory experiment.

The panel endorses continued development of technologies for neutrino observatories, which may lead to a high-statistics detection of cosmogenic neutrinos. A particularly challenging aspect is self-triggering using radio data alone at sites near human populations and thus radio backgrounds. As this field has many proposed techniques and the instrumental technology is developing rapidly, the choice of the observational approach for field implementations would preferably be determined through competitive peer review. Owing to the potentially enormous scientific return, it is critical that this direction must be aggressively pursued during the coming decade. Collaboration with astronomers using large arrays for radio astronomy is also encouraged.

Closely related to the above, technology development is needed to work toward dramatic improvements in ultra-high-energy cosmic ray sensitivity, as an order-of-magnitude expansion of existing arrays would require deployment over greater than tens of thousands of square kilometers. Such detectors could also be sensitive to ultra-high-energy gamma rays from the nearest sources of ultra-high-energy cosmic rays. Modest funding for small development efforts may be available through NSF PI programs, or the NSF MRI program, but significantly increased development opportunities are needed. Below the ultra-high-energy scale, there are a wide variety of successful or planned cosmic-ray experiments, and these are critical for testing the origins of Milky Way cosmic rays, finding the PeVatrons, and probing dark matter.

L.4.4.2 Scientific Opportunities for Gamma-Ray Observatories

NASA's Explorer Program could provide high-impact opportunities to conduct multi-messenger astronomy, either through missions targeted for that purpose or through the additional capabilities of missions targeted to other specific purposes. Continuing the current frequency of opportunities for the openly competed Explorer Program is highly desirable, and an increase in the maximum allowed cost for missions would be warranted. The Astrophysics Research and Analysis (APRA) program and the new Astrophysics Pioneers program are vital for opening new areas for gamma-ray astrophysics and enabling technical developments that will lead to future-generation observatories. The shorter implementation time for these programs is also important to support imminent multi-messenger and particle astrophysics advances and allow responsiveness to emerging discoveries.

The panel reviewed a rich array of concepts for new space-based gamma-ray observatories. These covered a range of scientific opportunities and made use of a range of techniques and mission scales,

representing significant progress made since 2010, in particular for the energy band from 100 keV to 100 MeV, which is markedly underdeveloped compared to lower- and higher-energy bands. In many cases, the technical readiness for concepts is already high and would allow significant observations in targeted capability areas. Examples of exciting scientific opportunities at the small scale include studies of the history of nucleosynthesis in the Milky Way, measurements of dynamic tomography of Type Ia supernovae, and measurements of gamma-ray polarization.

Several concepts were considered that would make excellent pathfinder studies. Small-mission opportunities in the next decade can also lead to future observatories that will provide additional leaps in the number and type of gamma-ray sources that can be studied in the higher-energy band from 100 MeV to 100 GeV. Continuing the current pace of opportunities for smaller missions will be critical for advancing observational progress in these new areas. Additionally, it will be necessary to support technology development to enable broadly capable next-generation high-energy space-based observatories. A vital piece of the development program for gamma-ray and cosmic-ray observatories is a well-supported balloon program. Development work in the next decade will be critical to attaining the capabilities that will be needed to support the breakthrough multi-messenger science possible in the 2030s.

The capabilities for wide-field detection of energetic phenomena and rapid multiwavelength follow-up provided by currently operating high-energy space missions, such as Fermi and Swift—which have finite lifetimes and no clear successors—have supported an extremely rich range of synergistic science discoveries involving gravitational waves, neutrinos, and very-high-energy gamma-ray observations. Until missions that exceed the current sensitivity and output become available, it is critical to continue support of these high-impact facilities.

L.5 RATIONALE FOR THE PROGRAM

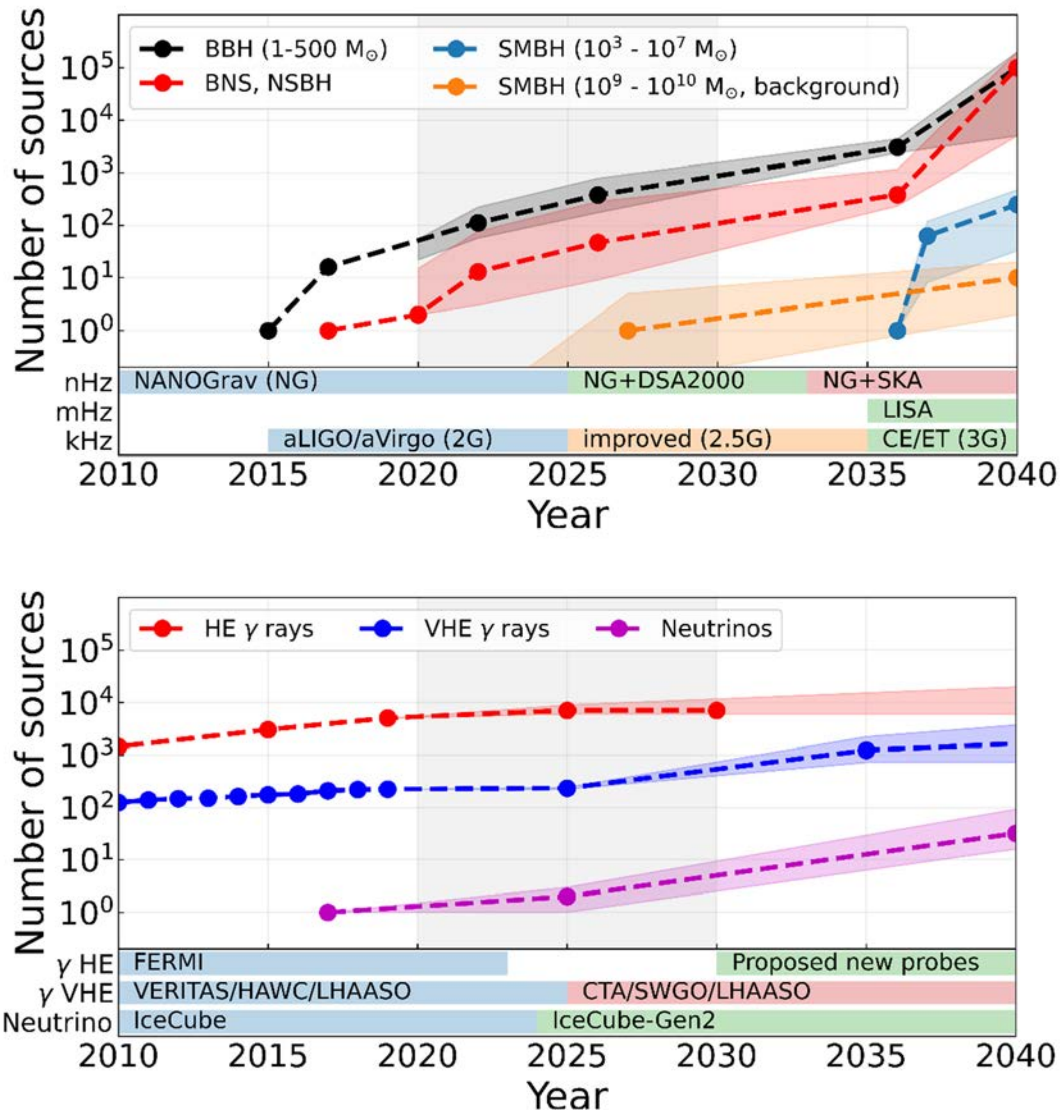
As regards the topics considered by the panel, Astro2020 is a special Decadal Survey. Compared to Astro2010, the scope of PAG's activities that have direct relevance to astronomy has dramatically increased owing to breakthrough discoveries. LIGO has detected dozens of gravitational-wave sources; IceCube has detected a bright, hard-spectrum diffuse background and one likely source; and gamma-ray observatories have played critical roles in leveraging those observations plus in making their own discoveries. In contrast, at the time of Astro2010, there were no direct detections of astrophysical sources of gravitational waves or very-high-energy neutrinos, and the impetus for multi-messenger astronomy was notional at best.

Astro2020 is also special compared to Astro2030: the 2020s are a critical time to act to maximize the returns of current and pending investments. Much of the science and technology in PAG's scope was developed in the United States, which excited worldwide interest, leading to huge investments abroad that take advantage of the early investments here. Action is needed now to maintain U.S. leadership and to develop the next stages of the science and technology. In some cases, it would be advisable for the United States to lead bold efforts; in others, it would be acceptable if it participated as a junior partner. But the United States must not cede its leadership in particle astrophysics and gravitation, which is of central and growing importance to both astronomy and physics.

The projects considered by the panel span a wide range of topics—gravitational waves, neutrinos, gamma rays, and cosmic rays—where continued investment in each area nurtures unique power to address multiple science-panel questions. *The first overall observation of the panel is that astronomy with new messengers is astronomy per se.* There is a high priority for observations of **extreme gravitators**, which include electromagnetically dark mergers of black holes, as well as of **extreme accelerators**, which include gamma-ray obscured sources of high-energy neutrinos. But this is not the whole story, as emphasized through the COEP Panel's discovery area of multi-messenger astronomy. *The second overall observation of the PAG Panel is that coordinating new-messenger capabilities is essential.* Without this coordination, we will not be able to fully understand multi-messenger sources like binary neutron star

mergers, cosmic-ray accelerators, a Milky Way supernova, and more. Multi-messenger observations are especially critical for rare, spectacular transients, where the opportunity for incredible insights depends on the completeness of the coverage. The panel has thus chosen to endorse key projects in multiple areas. Despite the number of projects, the overall costs are modest, and they would be partially funded from physics programs. Projects using individual new messengers could each lead to major discoveries in astronomy. Together, as multi-messenger astronomy, the prospects are even greater.

Figure L.4 summarizes how science capabilities would be enhanced in the 2020s and 2030s by the endorsed program, as well as the essential need for time and capability coordination between experiments to maximize the potential for multi-messenger astronomy. In addition to new observatories in the 2020s, research and development investments are needed to enable even larger scientific payoffs in the 2030s.



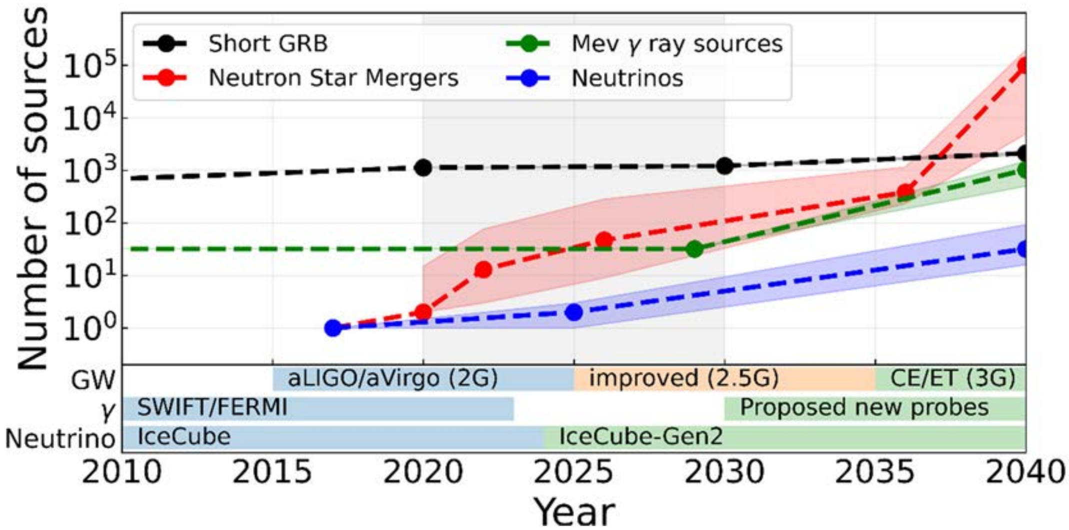


FIGURE L.4 Following Figure L.2, but in more detail, estimated projected cumulative numbers of detected sources over time possible with new investments. The three panels focus on gravitational-wave sources, neutrino and gamma-ray sources, and multi-messenger sources, respectively. As discussed in the text, even small numbers of new sources can have powerful impacts in terms of science results and in laying the groundwork for higher-statistics observations in the 2030s.

L.6 PROGRAMMATIC ISSUES

There are programmatic issues that need to be addressed to nurture the fields in the panel's scope. Some are cross-cutting, applying to the whole Astro2020 Decadal Survey.

Funding—The program above would fit within existing funding profiles for the NSF MREFC, NASA Probe-scale, and NSF Mid-scale programs. However, these programs are intensely competed among many fields. Further, several of the PAG-endorsed programs would be competing with each other in the NSF Mid-scale program, which makes it more difficult to enable a complementary suite of projects for multi-messenger astronomy. Another issue is that there needs to be a well-thought-out plan to fund operations of projects built by the MREFC program without negatively impacting grant funding in the corresponding research areas. NSF has suggested that the scope of the MREFC program could be changed to include initial funding for operations. The panel endorses this. Longer-term, the MREFC budget is not likely to be sufficient to support large projects such as Cosmic Explorer. For all the reasons above, it is vital to the field of astronomy to push for the budgets of all of these programs to be expanded while ensuring cost caps so that overruns for larger projects do not erode support for smaller projects and investigator-led programs.

For many projects within the scope of the panel, the Department of Energy (DOE) could be a good fit in terms of agency objectives, and DOE leadership or partnership would bring important topical expertise, plus critical experience in managing large facilities and collaborations. Astronomical observations with new messengers directly address fundamental questions in particle and nuclear physics, including the properties of neutrinos and particles beyond the standard model, the high-density equation of state and the production of the chemical elements, and many aspects of cosmology. Full or joint funding by DOE has been done successfully and could be repeated.

Given the wide range of participating observatories on the ground and in space, unprecedented coordination within and between agencies, and among international partners, is needed to establish a robust program to ensure that the right projects are operating at the same time to maximize the science opportunities. This requires increased coordination over multiple funding cycles and special care that

potential breakthrough projects do not fall between the cracks. The panel encourages enhanced communication between NASA, NSF, DOE, and international partners to guarantee science opportunities are not missed. In particular, the panel encourages that agencies develop a coherent long-range strategic plan that ensures coordinated coverage over the multiple wavelengths of multiple messengers, paying special attention to the timing for new instruments to come online and old instruments to be decommissioned.

Despite the necessary trend toward “big” science, avenues for funding small groups acting independently of the dominant trends are critical to innovating and laying the seeds for the science of subsequent decades. Several of the white papers relevant to the panel describe large, bold projects that would open new areas of observation space, such as localized cosmic ray sources, for example, but these projects are only at the early stages of conception. NASA and NSF support for technology development and small precursor missions is essential so that by the 2030s new capabilities will be available and ready. For example, the NASA Strategic Astrophysics Technology (SAT) program could support multi-messenger technology development. As the precision of large observatories improves, it will be increasingly important to also support measurements of fundamental atomic, nuclear, and particle physics processes needed to interpret the observations.

Theoretical and computational investigations play a key role in the advancement of knowledge by developing new laws, techniques, and ideas for experiments. For multi-messenger astronomy in particular, there is also a critical need to connect the results of different projects. However, by “following the science,” investigators may find themselves unable to fit within the boxes defined by the agencies, especially across the physics-astronomy boundary. In addition, many programs—for example, the NSF Astronomy program and the NASA Astrophysics Theory Program—are extremely oversubscribed. The panel encourages new, larger lines of support for theory critical to multi-messenger astronomy, particularly at the intersection of agency-defined traditional boundaries, plus a robust theory component to Guest-Investigator programs of future supported facilities and missions.

Operations Ecosystem—The success of multi-messenger astronomy relies on the coordinated efforts of astronomers spanning a wide range of groups, nations, and observational facilities. A robust “ecosystem” of observatories, some outside the direct purview of this Panel, is needed to localize and characterize the electromagnetic counterparts of gravitational-wave and neutrino sources. Given the wide fields of view of gravitational-wave and neutrino observatories, optical sky monitors that cover the whole sky are required. Although the Rubin Observatory's LSST project will provide broad, deep coverage in the Southern Hemisphere, complementary facilities are needed in the Northern Hemisphere, as are full-sky monitors (e.g., the All-Sky Automated Survey for Supernovae [ASAS-SN]) for optical transients too bright for the LSST project. In addition, optical spectroscopic facilities with flexible scheduling are required to confirm and characterize the huge numbers of photometric candidates and identify the source redshifts. As multi-messenger, time-domain astrophysics grows to become a larger part of astronomy, capabilities for multi-messenger follow-up become a fundamental consideration of data management/operations plans at the early stage of all mission concepts, even if the principal science drivers are not multi-messenger focused. It would be ill-advised for projects to eliminate critical capabilities—such as the ability to slew rapidly or to accept and issue real-time alerts—without explicit consideration of the cost and trade-offs associated with multi-messenger capabilities.

The timely release of public alerts is critical to maximizing the scientific gains by multi-messenger discoveries. As the number of discoveries by LIGO, IceCube, and other projects grow, the competition for precious multi-wavelength follow-up observations on large facilities will increase accordingly. To make efficient use of sparse resources, the panel endorses an open data policy, particularly regarding information—timing, sky position, and other decision-critical information—needed to coordinate and prioritize multi-messenger follow-up. This could be achieved by explicitly considering the timing and detailed information content of public data releases in evaluating mission and facility proposals. Requirements for timely public data releases will require dedicated funding for validating and documenting the releases.

A new level of coordination across scientific communities is needed. There are several possible elements. For transients, coordinated observations are critical, but this can be very difficult to arrange, owing to requiring separate proposals, often with time scales incompatible with each other and with a quick response. As recommended by the NASA Gravitational Wave—Electromagnetic Counterpart Task Force, there is a need for proposal calls that enable and prioritize joint observations. Once the data are in hand, the analysis of multi-messenger data is made more difficult if every data set comes in a different format and requires different tools. It would save a great deal of time for the community as a whole if there were adequate support to develop formats, tools, and alert standards that could be adopted by many projects. Likewise, there is a need for centralized, standardized ways of archiving and serving data. Last, it is important to build cross-project community ties to foster cooperation and innovation, including between astronomy and physics. The best solutions for these issues for particle astrophysics and gravitation are yet to be developed. Some aspects of the solutions may draw from the examples of the NSF-supported National Optical-Infrared Astronomy Research Laboratory (NOIRLab) and NASA's High Energy Astrophysics Science Research Center (HEASARC), which provide integrated platforms for multi-observatory science proposals, common data formats, and integrated analysis tools within their respective sub-disciplines. Last, the panel endorses the LISA Science Support Center as a mechanism to connect the U.S. and international communities; this may also be an example to other projects.

Culture—The fields within the scope of the panel span a wide range of scientific communities with disparate backgrounds and cultures. There is significant representation of people from physics, often from backgrounds in particle and nuclear physics, and funded through different mechanisms. Although such cultural differences are, on one hand, an obstacle that must be overcome to enable the vigorous research program laid out here, they also represent a great opportunity to bring fresh ideas and perspectives and to build a new field "from the ground up." New cultural work is needed to better connect physics and astronomy, as well as the sub-fields of gravitational-wave, neutrino, gamma-ray, and cosmic-ray astronomy, and to train students to cross these boundaries.

New cultural work is also needed to manage the increasingly large, international collaborations in ways that encourage the growth of communities that emphasize diversity, equity, and inclusion through proactive policies, mentoring, and accountability. The high visibility of particle astrophysics and gravitation, particularly among junior researchers, provides an opportunity to grow a joint community of outstanding vibrancy and diversity, one where individuals can have a big impact. Success will lead to new perspectives and discoveries, as well as societal benefits.

The responsibility for addressing these issues lies in multiple places. The funding agencies have a responsibility to take positive action to support traditionally excluded groups (race, gender, sexual orientation, and other minorities) in large collaborations, to give small grants to scientists whose work is outside of the boundaries or the traditions of the large collaborations, and to explicitly fund development at the intersections between physics and astronomy. Scientists themselves have the responsibility to engage in anti-racist, anti-sexist, and more general anti-discriminatory practices; to leverage the excitement of their fields to recruit underrepresented populations; and to nurture the next generation of scientists to develop interest and talent at the intersections between fields. The results will be transformational for both the community and the science.

M
Report of the Panel on Radio, Millimeter, and Submillimeter Observations from the Ground

M.1 EXECUTIVE SUMMARY

Observations at radio, millimeter, and submillimeter (RMS) wavelengths have played a critical role over the past decade in advancing our understanding of fundamental physics, cosmology, and the formation and evolution of cosmic structures on all scales (planets, stars, galaxies, and galaxy clusters). This record encompasses discoveries made by two types of RMS facilities: *experiments*, which are designed, built, and used by dedicated teams to address focused sets of science questions, and *observatories*, which offer diverse and flexible sets of observational capabilities to broad communities of astronomers, and can therefore address wide ranges of science questions. In the next decade, existing and new RMS facilities of both types are poised to make exciting discoveries in nearly all of the high-priority areas identified by the Astro2020 science panels. In this report, the Panel on Radio, Millimeter, and Submillimeter Observations from the Ground ("RMS panel") outlines the investments by U.S. federal agencies that would be most valuable for ensuring that this potential is realized.[1] This vision is balanced in terms of operational mode (experiment versus observatory) and scale (large versus medium versus small), and includes the following elements, listed in order of decreasing cost of construction per project:

- *Design, construction, and early operation of a large new observatory, the next generation Very Large Array (ngVLA).* This facility is conceived as an array of antennas distributed across North America, operating at frequencies from 1.2 to 116 GHz, and would replace two existing federally funded facilities—the Karl G. Jansky Very Large Array (JVLA) and the Very Long Baseline Array (VLBA). The ngVLA would provide dramatic improvements in the ability to detect and image faint astronomical signals at high angular resolution, enabling routine observations of cold gas flows inside distant galaxies, annular gaps produced by newly formed planets in the inner parts of protoplanetary disks, and features on the surfaces of nearby stars. With broad, flexible capabilities and science-ready data products accessible to a diverse community of users, the ngVLA would epitomize the strengths of observatory-mode science and enable discoveries in new areas that cannot currently be imagined.
- *Design, construction, and early operation of a large new "stage 4" experiment to study the cosmic microwave background (CMB), CMB-S4.* This facility would build on the achievements of previous (second- and third-generation) CMB experiments in Antarctica and Chile, deploying a suite of small- and large-aperture telescopes equipped with unprecedented numbers of detectors spanning many bands across a decade in frequency. Working together, the CMB-S4 telescopes would conduct two complementary surveys probing the afterglow of the Big Bang, placing unprecedentedly tight constraints on the strength of primordial gravitational waves and the contribution of light particles to the density of matter in the Universe. In addition to its unique ability to address longstanding questions of cosmology and

[1] See Appendix A for the overall Astro2020 statement of task, the set of panel descriptions that define the panels' tasks, and for additional instructions given to the panels by the steering committee.

fundamental physics, CMB-S4 would be poised to shed light on the growth of cosmic structures (in particular, formation of the first galaxy clusters) and the properties of explosive transients in a new frequency regime.

- *Significant funding to support mid-scale projects.* In the past decade, NSF mid-scale programs offering up to $70 million of funding per project have provided essential support to a number of cutting-edge RMS initiatives. While future funding decisions would be determined by competitive proposal calls, the RMS panel has identified four key areas in which outstanding opportunities exist for new mid-scale facilities to address compelling science questions. Listed in order of nearest to most distant observational target(s), these are:
  - Broadband, high-cadence, spectropolarimetric imaging of the Sun, to trace flares, shocks, and coronal mass ejections, and understand the drivers of space weather;
  - High-resolution imaging of jets driven by supermassive black holes in the centers of galaxies, to determine how such jets are launched and powered;
  - Surveying the static and time-variable radio sky with an innovative new "radio camera," to address a wealth of science questions using statistical samples of star-forming galaxies and fast radio bursts; and
  - Mapping the evolution of neutral atomic hydrogen (HI) gas in the very early Universe, at epochs before galaxies and black holes were sufficiently numerous to ionize it.

Calls for mid-scale funding issued on a regular basis over the next decade would accommodate a range of projects reaching readiness on different time scales, and would enable agile and cost-effective approaches to addressing new science opportunities.

- *Ongoing support for three key capabilities—long-term timing of pulsars, development of new instrumentation (including software), and mitigation of radio frequency interference (RFI)—that are not tied to single facilities.* The precision timing of pulsars encompasses both individual systems and large networks of objects, with the latter aimed at enabling the detection of low-frequency gravitational waves, and requiring ongoing searches to expand existing networks. These efforts require continued access to substantial observing time on the Arecibo telescope and the Green Bank Telescope (GBT), with the ngVLA and a future mid-scale facility potentially contributing as well near the end of the decade. Given the critical importance of Arecibo and the GBT for pulsar timing, and their value in addressing other high-priority science questions, continued federal (and, if available, state) funding to support healthy fractions of "open time" scientific observations at these observatories would be very important. More broadly, to maintain the capacity to build and exploit innovative new RMS instrumentation (including software) and train the next generation of instrument builders, dedicated federal funding (e.g., via the NSF Advanced Technologies and Instrumentation program) remains critical, as does the need for platforms where new instruments can be deployed. The rapid increase in RFI from terrestrial sources and satellite constellations poses a severe threat to radio astronomy. The alleviation of this threat will require increased support for RFI protection and mitigation for all new and existing ground-based facilities operating at RMS wavelengths.

The RMS panel recognizes its role as helping to populate a menu of options from which the Astro2020 steering committee will choose in recommending an ambitious program for the next decade. As an input to this process, the RMS panel has also identified three top-level principles governing its overall vision. First, it would be important for facility operations budgets over the next decade to include full support for the U.S. share of the Atacama Large Millimeter/submillimeter Array (ALMA), a hugely productive and scientifically vibrant observatory that has set a new standard for how an RMS facility can serve the entire astronomical community. Second, RMS science will flourish best with a balanced

program of investments at large, medium, and small cost scales, where “small” investments include individual investigator grants that can support the training of graduate students and postdoctoral researchers—and with the level of investment at any one cost scale not becoming so large that balanced investment at a smaller cost scale is precluded. Third, in the construction and operation of RMS facilities, the astronomical community needs to engage constructively, respectfully, and substantively with stakeholders from outside that community in managing cultural and environmental concerns.

BOX M.1

On December 1, 2020, following earlier support cable failure in August and November, the Arecibo telescope suffered a catastrophic collapse. The timing of this event relative to the progress of the Astro2020 survey has precluded any detailed response in this report, although it is clear that the loss of Arecibo’s capabilities will significantly impact the ability of the U.S. astronomy community to address high-priority Astro2020 science questions. To illustrate the full breadth and depth of these impacts, the RMS panel has left the portions of this report pertaining to Arecibo essentially unchanged from what was submitted to the Astro2020 steering committee in July 2020. The panel affirms that the impacts of Arecibo’s loss can only be mitigated by the investment of additional observing time on existing and/or new facilities.

M.2 THE RMS LANDSCAPE IN 2020

M.2.1 Looking Back

In August 1931 in Holmdel, New Jersey, using a 100-foot-long antenna mounted on four Model T tires, American physicist and engineer Karl Jansky detected radio waves originating in the center of the Milky Way. Following developments in radar in World War II, the discipline of RMS astronomy grew and ramified to produce a stunning range of technological advances and scientific discoveries. RMS telescopes can observe in isolation, as single dishes sensitive to the diffuse emission from interstellar gas clouds and the faint pulses of spinning neutron stars, or in concert, as arrays (sometimes continental or intercontinental in scale) producing exquisitely sharp images of galaxies, black holes, and protoplanetary disks. RMS detectors on these telescopes can distinguish the signatures of specific atoms and molecules from those of thermal plasmas that glow because they are warm, and in turn from the nonthermal emission produced by charged particles accelerating in strong magnetic fields. Modern astronomers can leverage these RMS technologies and techniques to study phenomena ranging from explosive events on the surface of the Sun to tiny ripples in the cosmic microwave background that represents the afterglow of the Big Bang.

An important dimension of RMS astronomy is the fact that groundbreaking discoveries are regularly made by facilities that operate in two different modes. Experiments are conceived, constructed, and exploited by dedicated teams to address focused sets of science questions, with design parameters optimized to deliver the best possible performance in addressing those questions. Observatories are designed and built to be broadly capable, and thus able to address wide ranges of science questions — often much wider than the original designers and builders imagined. An observatory achieves its full potential by eliciting the most creative and ambitious ideas from the broadest possible community of astronomers; this consideration explains why the National Radio Astronomy Observatory (NRAO), the Green Bank Observatory (GBO), and the Arecibo Observatory have made it part of their core mission to expand their user communities through support, outreach, and training activities. Experiments and

observatories are essential complements to each other in the progress of science. Figure M.1 shows examples of discoveries made since 2010 by observatories, experiments, and a combination of the two.

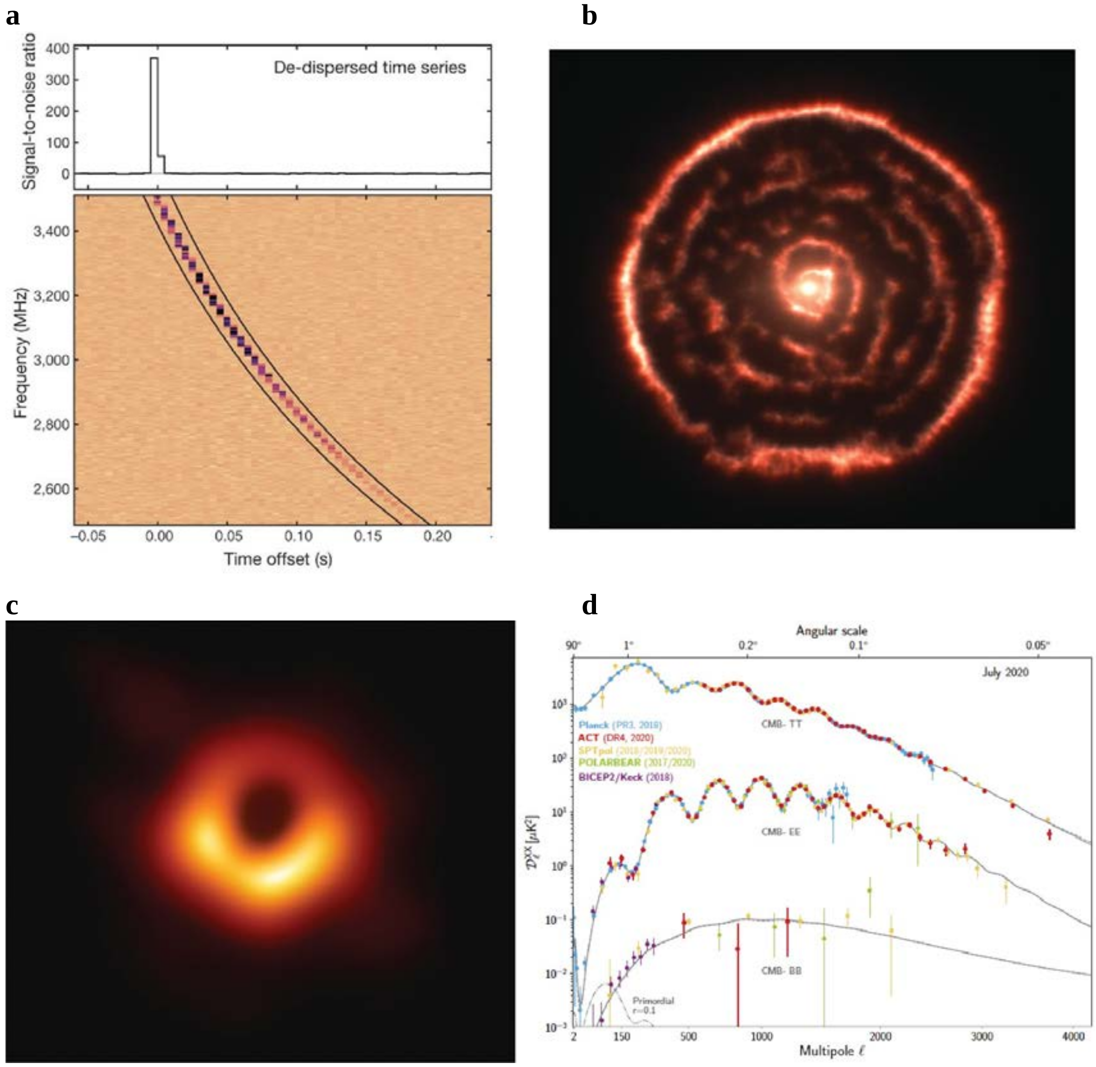


FIGURE M.1 (a) First localization of a fast radio burst (FRB) by Chatterjee et al. (2017), who used the JVLA to pinpoint the location of a burst initially detected at Arecibo. (b) CO emission from the circumstellar envelope and shell around the evolved star R Sculptoris, imaged by Maercker et al. (2012) using ALMA. (c) Image of the supermassive black hole at the center of the galaxy M87, obtained by the Event Horizon Telescope observatories in conjunction with ALMA (Event Horizon Telescope Collaboration et al., 2019). (d) Temperature anisotropy and polarization measurements made by second-generation CMB experiments, including the ACTPol, South Pole Telescope (SPT), POLARBEAR, and BICEP2/Keck experiments (adapted from Choi et al., 2020). SOURCE: (a) S. Chatterjee, C.J. Law, R.S. Wharton, S. Burke-Spolaor, J.W.T. Hessels, G.C. Bower, J.M. Cordes, et al., 2017, A direct localization of a fast radio burst and its host, Nature 541(7635):58–61. (b) ALMA (ESO/NAOJ/NRAO) and M. Maercker, S. Mohamed, W.H.T. Vlemmings, S. Ramstedt, M.A.T. Groenewegen, E. Humphreys, F. Kerschbaum, et al., 2012, Unexpectedly large mass loss during the thermal pulse cycle of the red giant star R Sculptoris, Nature 490(7419):232–234. Reproduced with permission. (c) Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A. Baczko, et al., 2019, First M87 event horizon telescope results, Nature 875(1):L1. Courtesy of The Event Horizon Telescope Collaboration. (d) NSF Adapted from S. Choi, M. Hasselfield, S.P. Ho, B. Koopman, M. Lungu, M.H. Abitbol, G.E. Addison, et al., 2020, The Atacama Cosmology Telescope: a measurement of the Cosmic Microwave Background power spectra at 98 and 150 GHz, Journal of Cosmology and Astroparticle Physics 12(045), 045 © IOP Publishing Ltd and Sissa Medialab. NSF Reproduced by permission of IOP Publishing. NSF All rights reserved. doi:10.1088/1475-7516/2020/12/045.

Scientifically, highlights from the past decade of RMS observations have extended across all areas of astronomy. Multiple experiments have detected lensing B-mode signatures in the CMB, and the CMB-galaxy lensing cross power spectrum. Fast radio bursts (FRBs) have gone from unconfirmed curiosities to daily events that show promise as cosmological tools for probing the distribution of ionized gas across the Universe, as new facilities yield increasing numbers of detections and localizations. Centimeter wavelength measurements of neutron star masses more than twice that of the Sun have provided the strongest constraint yet on the super-nuclear equation of state, while radio observations of a binary neutron star merger have offered robust constraints on the geometry of the explosion and the expansion of the Universe. At (sub)millimeter wavelengths, ALMA's unparalleled sensitivity and resolution have allowed unprecedentedly detailed mapping of the molecular gas and dust in nearby galaxies (from which stars form), and around young stars (from which planets form). The striking ALMA image of the disk surrounding the young star HL Tau revealed exquisitely detailed structure—nested rings and gaps thought to be created by embedded planets—which, in conjunction with other studies, strongly suggests that planet formation is more extensive, more diverse, and earlier-starting than anticipated. ALMA spectral line observations have revealed the kinematics of disks around forming stars and black holes, thereby enabling measurements of their masses, and have probed the gas mass reservoirs in high-redshift galaxies, revealing the factors that drive the cosmic star formation history. Detailed images of gravitationally lensed galaxies have demonstrated the potential to detect and measure the masses of dark matter subhalos, while closer to home, spectacular ALMA images of the solar chromosphere have provided essential data for studying the outer layers of the Sun.

Technologically, emerging trends from the past decade have enabled faster, low-cost prototyping and development of new facilities. At (centi)meter wavelengths, new developments have largely followed the trajectory of commercial products and are drawn from four categories: low-cost, low-noise amplifiers at increasingly high frequencies, and low-cost, high-bandwidth digital samplers (both driven by telecommunications needs); continued expansion of computing and especially highly parallelizable graphics processing units (driven by data science/gaming needs); and high-bandwidth network switches (driven by telecommunications and high-performance computing needs). Taken together, these devices make possible inexpensive, sensitive, wide-bandwidth radio receivers, correlators, and complex real-time data processing, such that instruments can be rapidly prototyped using commercial off-the-shelf components and software-defined radio tools. At (sub)millimeter wavelengths, the commercialization of cryogenics and the development of large arrays of superconducting detectors have dramatically increased the numbers of detectors that can be deployed on wide-field telescopes, allowing sky-statistics-limited (rather than detector-limited) analyses to be conducted for the first time.

Programmatically, the most dramatic development of the past decade has been the emergence of ALMA as a facility that engages the full (in terms of both wavelength and geography) astronomical community. High demand for observing time by a user base much broader than the traditional RMS community, a large and growing stream of impressive published results, and a well-defined pathway to future upgrades using dedicated development funding have set ALMA apart from previous RMS observatories. Two key factors in ALMA's success are its superb imaging performance, and its investment in pipeline development and provision of high-quality data products (through its archive) to users who are nonexperts and/or have limited computing capacity at their home institutions. Both factors will also be relevant to the success of future RMS observatories.

M.2.2 Looking Forward to the Next Decade

In looking ahead to a decade (2022–2032) that will include the centenary of Karl Jansky's pioneering discovery, the RMS panel relied on inputs from a variety of sources. Foremost among these was an impressive suite of white papers—innovative, ambitious, and wide-ranging—that laid out projects and priorities for the next 10 years and beyond. In reviewing these white papers and engaging with the

teams who submitted them, the panel was guided by the high-priority science questions and discovery areas identified by the six Astro2020 science panels. In most cases, the RMS panel concurred with the science panels' own assessments of where and how RMS observational capabilities could help address these questions; in a few cases, opportunities for impact beyond those identified by the science panels were factored into the RMS panel's analysis. In line with the additional guidance provided to the Astro2020 survey, the RMS panel also considered the findings and recommendations of the National Academies reports, *Solar and Space Physics: A Science for a Technological Society* (2012), *Exoplanet Science Strategy* (2018), and *An Astrobiology Strategy for the Search for Life in the Universe* (2019).

To characterize the ability of existing and proposed RMS facilities to address high-priority science questions and discovery areas (in concert with multi-wavelength and multi-messenger observations, theoretical work, and laboratory investigations), the RMS panel developed a scoring rubric with three categories. These distinguish areas where (1) a facility would make a contribution in addressing a science question (or any of its sub-questions) that would be irreplaceable and unique relative to other facilities with U.S. community access; (2) a facility would make a very significant contribution in addressing a science question but would not be sufficient to address that question by itself (e.g., in the absence of observations at other wavelengths); and (3) a facility would have an impact in addressing a science question, but would be one of several facilities playing supporting roles. To assess the risks and costs of large projects (i.e., those requiring > $70 million of federal funding), the RMS panel considered presentations from and extensive documentation provided by the respective project teams, while also making use of independent analyses by The Aerospace Corporation in the context of a Technical, Risk, and Cost Evaluation (TRACE) process (see Appendix O). The panel adopted a "hybrid" approach for these assessments, producing syntheses of risk registers and schedule and cost estimates from the project teams and from the TRACE analyses that are informed by panel expertise and represent the panel's best judgments.

As a result of these deliberations, the RMS panel arrived at a set of four priorities for new or enhanced federal investment over the next decade. In order from highest to lowest construction cost per project, these investments are:

- Design, construction, and early operation of a large new observatory, the next generation Very Large Array (ngVLA);
- Design, construction, and early operation of a large new "stage 4" experiment to study the CMB, CMB-S4;
- Significant funding to support mid-scale projects with costs of up to (at least) $70 million—where exciting opportunities exist in areas that include (but are not limited to) broadband, high cadence, spectropolarimetric imaging of the Sun, high-resolution imaging of jets driven by supermassive black holes, surveying the static and time-variable radio sky, and mapping the evolution of neutral atomic hydrogen (HI) gas in the very early Universe—and
- Ongoing support for three key capabilities, the long-term timing of pulsars (which will require continued support for operation of the Arecibo telescope and the GBT), the development of new instrumentation (including software), and the mitigation of RFI.

The scientific motivations for these investments are discussed in detail in the relevant sections below, as are risk, cost, and programmatic issues for the ngVLA and CMB-S4. Table M.1 provides a concise visual representation of the RMS panel's scoring of all relevant existing and proposed facilities against the Astro2020 high-priority science questions, with darker shades indicating areas where facilities can make more substantial contributions. The report concludes with suggestions of guiding principles for achieving balance (existing versus future facilities, large versus medium versus small cost scales, astronomers versus other stakeholders) within the RMS portfolio.

TABLE M.1 High-Priority Science Questions Versus RMS Facilities

Science Frontier Panel Questions / Discovery Areas	ngVLA	CMB-S4	MSO1	MSO2	MSO3	MSO4	Arecibo	GBT	ALMA	VLBA	VLA
Panel on the Interstellar Medium, Star and Planet Formation (ISM)											
1) How do star-forming structures arise from, and interact with, the diffuse ISM?											
2) What regulates the structure and motions within molecular clouds?											
3) How does gas flow from parsec scales down to protostars and their disks?											
4) Is planet formation fast or slow?											
D) Detecting and characterizing forming planets											
Panel on Exoplanets, Astrobiology and the Solar System (EAS)											
1) What is the range of planetary system architectures and is the configuration of the solar system common?											
2) What are the properties of individual planets, and which processes lead to planetary diversity?											
3) How do habitable environments arise and evolve within the context of their planetary systems?											
4) How can signs of life be identified and interpreted in the context of their planetary environments?											
D) The search for life on exoplanets											
Panel on Stars, the Sun and Stellar Populations (SSSP)											
1) What are the most extreme stars and stellar populations?											
2) How does multiplicity affect the way a star lives and dies?											
3) What would stars look like if we could view them like we do the Sun?											
4) How do the Sun and other stars create space weather?											
D) "Industrial Scale" Spectroscopy											
Panel on Compact Objects and Energetic Phenomena (COEP)											
1) What are the mass and spin distributions of neutron stars and stellar black holes?											
2) What powers the diversity of explosive phenomena across the electromagnetic spectrum?											
3) Why do some compact objects eject material in nearly-light-speed jets, and what is that material made of?											
4) What seeds supermassive black holes and how do they grow?											
D) Transforming our view of the Universe by combining new information from light, particles, and gravitational waves											
Panel on Galaxies (GAL)											
1) How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?											
2) How do gas, metals, and dust flow into, through, and out of galaxies?											
3) How do supermassive black holes form, and how is their growth coupled to the evolution of their host galaxies?											
4) How do the histories of galaxies and their dark matter halos shape their observable properties?											
D) Mapping the circumgalactic medium and intergalactic medium in emission											
Panel on Cosmology (COS)											
1) What set the Hot Big Bang in motion?											
2) What are the properties of dark matter and the dark sector?											
3) What physics drives the cosmic expansion and the large-scale evolution of the Universe?											
4) How will measurements of gravitational waves reshape our cosmological view?											
Discovery Area: The Dark Ages as a cosmological probe											

NOTE: Top row indicates future large facilities (dark yellow), future mid-scale opportunities (MSO, light yellow), existing facilities that would continue to operate (blue), and existing facilities that would be replaced by the ngVLA (pink). Columns MSO1 through MSO4 refer to opportunities in the areas of broadband solar imaging, high-resolution imaging of jets, surveys of the static and time-variable radio sky, and HI in the early Universe, respectively. Rows list science questions and discovery areas, which are identified by panel acronym (ISM, EAS, SSSP, COEP, GAL, COS) and number (1, 2, 3, 4, D) elsewhere in the report. Each cell is shaded to indicate the role of the facility in addressing the topic. Dark green indicates the facility is irreplaceable (at any wavelength) for addressing one or more sub-questions within a topic, and unique relative to other facilities with U.S. community access; medium green indicates the facility is essential, but not sufficient to address a topic by itself; light green indicates the facility is one of many with supporting roles in addressing a topic.

M.3 A LARGE NEW OBSERVATORY: THE NGVLA

M.3.1 Introduction

Since its completion in 1980, the JVLA has been an outstandingly versatile and productive facility for advancing knowledge about the Universe at centimeter wavelengths. The JVLA identified the first Milky Way "microquasar" driving jets of energetic particles at close to the speed of light; determined the location of the black hole at the center of the Milky Way; discovered the first complete "Einstein ring" produced by gravitational lensing of a distant galaxy; and detected the first radio-wavelength counterpart to an explosive gamma-ray burst (GRB) event. In 2012, completion of the Expanded VLA (EVLA) upgrade yielded a more capable facility, with more sensitive receivers covering a wider range of frequencies, and a correlator able to process wider bandwidths. Even after these improvements, however, the JVLA has been hampered by the surface accuracies of its antennas, which limit its performance at the higher frequencies where emission from thermal processes is strongest, and by the number and allowable configurations of those antennas, which limit the quality of the images it can produce. Similarly, since coming online in 1993, the VLBA has blazed new trails in measuring the Hubble constant (via observations of water megamasers) and revealing the structure of the Milky Way. It too has been upgraded in bandwidth and frequency coverage, but its limited sensitivity (particularly at higher frequencies) constrains users' ability to take full advantage of its superb angular resolution. As reflected in the Astro2020 science panel reports (see also Table M.1), a wealth of discovery opportunities would be within the grasp of a centimeter-wavelength successor to the JVLA and VLBA that offered an order of magnitude improvement in sensitivity, and the ability to image sources on scales of arcminutes to fractions of a milliarcsecond (as appropriate for their surface brightness on those scales) across two decades in frequency. The RMS panel supports the funding of such a next generation Very Large Array (ngVLA) in order to realize this scientific potential.

M.3.2 Science Case

The ngVLA design concept (see below) has been optimized through extensive community engagement to deliver on five key science goals (McKinnon et al., 2019)[2] that are broadly aligned with the Astro2020 high-priority science questions. This section highlights areas in which the ngVLA's capabilities would allow it to make extraordinary contributions to addressing those questions, grouped by the RMS panel according to three broad science themes.

[2] M. McKinnon, A. Beasley, E. Murphy, R. Selina, R. Farnsworth, and A. Walter, 2019, ngVLA: The next generation very large array, white paper submitted to the Astro2020 Decadal Survey.

M.3.2.1 Stars and Planetary Systems

The births, lives, and deaths of stars, and the formation and evolution of planets in orbit around them, represent an important focus for RMS observations. Understanding the origin and prevalence of high-density structures within larger molecular clouds, and the role these structures play in the formation of stellar nurseries and the birth of stars, is critical for elucidating the first stages of the stellar life cycle. Young stellar systems can be explored for evidence of young planets, to provide insights on the environments in which planets form and the speed with which they do so. Astronomers are now on the verge of being able to understand the architectures of planetary systems, the means by which planets migrate to different locations within these systems, and the origin, evolution, and prevalence of habitable environments. Understanding stellar activity at all phases of a star's life is important for predicting potential impacts on planets in its habitable zone, while near the end of its life, the process of mass loss and its connection to stellar death become increasingly relevant. The following paragraphs highlight ngVLA capabilities that would address high-priority questions identified by the ISM, EAS, and SSSP science panels.

The ngVLA's combination of high spatial resolution and centimeter wavelength continuum sensitivity would resolve protoplanetary disks on scales more than 20 times finer than ALMA, potentially capturing images of planet formation in action as young planets clear gaps in the disk. The inner regions of these disks, within a few astronomical units of the central star, are often too opaque to be studied at the shorter ALMA wavelengths. The ngVLA would image circumstellar disks in hundreds of protoplanetary systems with sufficient resolution (~5 milliarcseconds, corresponding to, for example, 0.6–0.9 AU for the Taurus molecular cloud) to measure inner-disk surface density perturbations caused by young, forming super-Earths. These disk substructures would offer important insights on the process of planet formation (ISM-4). The short orbital periods of planets close to host stars could potentially be tracked through very sensitive ngVLA imaging of disk structures on time scales as short as a few weeks. High-resolution centimeter and millimeter studies of large samples of disks would help develop a census of planetary system architectures (e.g., distribution, mass, orbital radii) in mature systems to compare to those in newly forming protoplanetary disks (EAS-1), thereby helping researchers to understand the diversity of planetary systems, how our own solar system formed, and how unique our solar system is (EAS-2).

The sensitivity, resolution, and imaging fidelity of the ngVLA would be used to map the physical conditions and gas motions within star-forming cores. Centimeter wavelengths are particularly critical for tracing high-mass star formation in deeply embedded environments, as well as in the densest portions of infalling low-mass stellar cores, where sub-arcsecond resolution spectral studies would help constrain protostellar masses and trace collapse motions (ISM-3). Sensitive spectral line observations across centimeter (e.g., NH_3, deuterated molecules) and millimeter (e.g., N_2H^+, CO isotopologues) wavelengths at high spatial (<0.1 pc) and velocity (<0.1 km/s) resolution would provide a census of dense gas as a function of environment within our own Milky Way galaxy. The sensitivity of the ngVLA across the centimeter and millimeter bands would further enable sub-arcsecond spectroscopic studies of large samples of galaxies in multiple species (HCN, HCO^+, CO isotopologues, and different excitation lines), to trace the efficiency of cloud collapse into forming stars (ISM-2). ngVLA mapping of the neutral gas in nearby galaxies on the scale of individual star-forming clouds would make it possible to trace gas flows through the crucial atomic to molecular phase transition within the interstellar medium, as gas moves from being potential fuel for star formation toward the brink of actually forming stars (ISM-1).

Over half of the dust and heavy elements in the interstellar medium originate in the winds and outflows of dying stars known as red giants. One of the greatest challenges in understanding the physics, geometries, and time scales of these winds is that the atmospheres of evolved stars are governed by the interplay of complex physical processes, including pulsations, shocks, convection, magnetic fields, and the formation of dust and molecules. Studying these phenomena demands exquisite spatial resolution (corresponding to a small fraction of a stellar radius), coupled with the ability to monitor time-variable behaviors. The wavelength range covered by the ngVLA would probe the regions of red giant atmospheres beyond ~2 optical radii (in both spectral lines and continuum) where stellar winds are

launched and accelerated. The ngVLA's collecting area on long (up to continental) baselines would provide the combined sensitivity and ultra-high resolution (as fine as ~70 μas) needed to probe the spatially resolved atmospheric dynamics, magnetic fields, brightness temperatures, and surface features (e.g., spots and convective cells) of hundreds of red giants and follow their evolution over time (SSSP-3).

The ngVLA would have the power to resolve over 10,000 stars within the Milky Way, including the radio surfaces and extended atmospheres of hundreds of main sequence stars. Observables include both the thermal emission from their photospheres and chromospheres, and coherent nonthermal and incoherent gyrosynchroton emission resulting from magnetic activity and stellar coronae. The ngVLA would be able to resolve the crucial zone where the physical processes driving stellar activity manifest themselves, and its ultra-wide bandwidths would enable the dynamic spectro-imaging and spectropolarimetry necessary to study the evolution of flares, coronal mass ejections, and other active and magnetically driven phenomena (SSSP-1), including those in active binaries (SSSP-2). Such studies of stellar activity have taken on a heightened importance because of their relevance to space weather in extrasolar planetary systems, which may impact the development of life (SSSP-4). Measurements with the ngVLA's continental baselines would also be able to trace binary orbits (including those of ultracool dwarfs), thereby enabling direct mass measurements (SSSP-1).

M.3.2.2 Black Holes and Galaxies

Supermassive black holes (SMBHs) are both physically and phenomenologically central to their host galaxies. In their vicinity, gravitational potential energy of infalling material and spin energy are converted to copious electromagnetic radiation and powerful jets that extend hundreds to thousands of light years from their bases. The resulting radiative and mechanical feedback can have long-lasting impacts, including regulation of star formation in the host galaxies, although the transfer of energy from the base of a jet to the surrounding medium is still poorly understood. From a population perspective, detailed understanding of the role that black holes play in star formation and the evolution of their host galaxies will require a census of black hole growth through merging and accretion over cosmic time. At lower masses, spectacular gravitational wave detections have revealed the mergers of individual stellar-mass black holes forming more massive black holes, yet the extent of a population of intermediate-mass black holes, between the extremes of SMBHs and stellar-mass black holes, remains uncertain, and questions remain concerning the masses and spins of binary black holes prior to merger and how those properties map to stellar progenitors. This subsection highlights how the ngVLA would address key open questions on black holes and galaxies as identified by the COEP and GAL science panels.

The ground-breaking Event Horizon Telescope (EHT) observation of the base of the relativistic jet in M87 (see Figure M.1c) would be complemented by the ngVLA's exceptional ability to trace details of the structure and acceleration of relativistic particles along the full lengths of that and many other jets, from scales of a few parsecs to hundreds of kiloparsecs (well beyond the 0.01 pc region probed by EHT). The continental (~9000 km) baselines of the ngVLA, together with its full polarization capability, would provide the sub-milliarcsecond imaging needed from centimeter to millimeter wavelengths to trace details of the jet inclination, lateral structure, magnetic field strength, and variation of the Lorentz factor away from the launch point region in many black hole systems. These studies would reveal the composition of the jets, how particles are accelerated, and how the jet parameters vary with distance from where the jets are launched (COEP-3).

Beyond the study of relativistic jet properties, the ngVLA's resolution, sensitivity, and imaging fidelity across the centimeter to few-millimeter band would be critical in searches for the elusive accretion signatures of intermediate-mass black holes, and in efforts to develop a census of binary black holes to determine the role of mergers in the formation of supermassive black holes (COEP-4). Sensitive, high-resolution ngVLA spectral line studies of molecular (low-*J* CO) emission would reveal hidden details of the interaction of relativistic jets with their surrounding interstellar media (GAL-3).

Galaxies themselves often contain vast reservoirs of fuel for star formation, and these reservoirs are strongly influenced by feedback from stellar winds and supermassive black holes. Studies of high-redshift systems reveal that molecular gas broadly traces the cosmic star formation rate history of galaxies. The ngVLA would enable detailed measurements of the masses and kinematics of molecular gas clumps on sub-kiloparsec scales in typical (Milky Way-like) galaxies out to beyond $z \sim 2$, when the Universe was one quarter of its current age. These observations would be coupled with ngVLA studies of small-scale feedback, which is known to regulate accretion and reduce star formation efficiency (GAL-2). On larger scales, individual galaxies are embedded within a circumgalactic and intergalactic medium that must be accounted for in our understanding of the star formation fuel reservoirs of individual systems. The ngVLA would enable resolved spectral study of this surrounding material on scales of kiloparsecs down to hundreds of parsecs (GAL-D). ngVLA observations of the dense environment near the center of our own Milky Way, including the Central Molecular Zone and the Circumnuclear Ring, could reveal details of the energetics, motions, and physical characteristics of gas that serves as a template for understanding distant galaxies that cannot be observed at such high spatial resolution (GAL-4).

M.3.2.3 Transient Sources and the Explosive Universe

The study of transients cuts across both astrophysics and cosmology. These brief, energetic events can trace stellar deaths, which drive the chemical enrichment of their surroundings and lead to the formation of neutron stars and black holes. Signatures of the binary neutron star merger GW170817 were detected across the electromagnetic spectrum after the initial gravitational wave signature of coalescence, beginning with the burst of gamma-rays two seconds later, followed by an optical counterpart within 11 hours — and a JVLA radio detection of an emerging relativistic jet 16 days after the detection of gravitational waves. Such observations have opened a new era of multi-messenger astronomy, where studies of individual events are combined across the electromagnetic spectrum to build a detailed picture of the engine that powers the diversity of transients. Transients can be used as “standard candles” to trace the acceleration of the Universe, or as probes of the “missing baryon” content of the Universe as recently undertaken for a sample of well localized FRBs. The COEP and COS science panels have identified a number of questions, discussed below, which require the sensitivity, resolution, and imaging fidelity of a new centimeter/millimeter observatory such as the ngVLA.

With its microJansky flux sensitivity (owing to improved receivers and increased collecting area) and the sub-milliarcsecond resolution provided by its continental baselines, the ngVLA would be ideally suited to characterize the energy sources driving explosive transient events. It would map the energy distribution of the explosive ejecta driven outward by the ignition of the transients, search for newly launched relativistic jets, and constrain the development of pulsar wind nebula-like emission. These observations would transform our understanding of explosive phenomena in the Universe (COEP-2).

Monitoring of compact binary mergers using the high resolution and excellent imaging fidelity of the ngVLA would move well beyond the excellent initial studies of GW170817 with the JVLA and the High Sensitivity Array (VLBA supplemented by the JVLA and GBT), which discovered and tracked the radio afterglow of an off-axis relativistic jet driven by the merger. ngVLA monitoring of these systems would allow early detection of newly formed relativistic jets and study of jet evolution when present. As part of a larger multi-messenger study, these new observations would permit detailed mapping of the initial merger conditions to the energetic impact on the local environment (COEP-D). Increased sensitivity of the gravitational wave network over the next decade will yield large samples of events that would require sensitive, high-resolution radio follow-up with an instrument such as the ngVLA, which would search for and image emission from electromagnetic counterparts. These observations would represent a critical step in building a large sample of standard sirens to independently probe the cosmic distance scale (COS-4).

M.3.3 Design Concept

At the broadest level, the vision for the ngVLA entails an order of magnitude increase in capabilities over those of existing facilities. Both the ngVLA project's own key science goals and the high-priority questions identified by the Astro2020 science panels demand coverage of broad and continuous frequency ranges between 1.2 and 116 GHz, velocity resolution as fine as 100 m/s, sub-milliarcsecond angular resolution, and high-fidelity imaging capabilities on scales from milliarcseconds to arcminutes. The technical realization that satisfies these specifications consists of 244 reflector antennas of 18m diameter and 19 reflector antennas of 6m diameter, all at fixed locations. The Main Array (MA) and Short Baseline Array (SBA) would have 214×18 m and 19×6 m antennas, respectively, centered on the current JVLA site in New Mexico but distributed on baselines ranging from meters to ~1000 km across the southwestern United States and into Mexico. This configuration would allow for the sampling of a broad range of spatial scales (from arcminutes to milliarcseconds). A Long Baseline Array (LBA) of 30×18m antennas would be located in ten clusters (mostly at existing VLBA sites), providing continental-scale baselines and sub-milliarcsecond resolution. All antennas would be connected by optical fiber to a single flexible signal processing center, allowing for real-time correlation of all antennas simultaneously and operation in subarrays. Each antenna would feed a suite of cryogenically cooled receivers allowing operation from 1.2 to 116 GHz (except for the 50–70 GHz range where the atmosphere is opaque), which would provide access to the HI and CO(1–0) emission lines at $z = 0$, and with spectral resolution better than ~0.1 km/s. An ambitious software and archive effort is projected to allow science-ready data products to be generated promptly and shared with users via server-side visualization and analysis platforms, thereby enhancing prospects for archival research, and lowering barriers to astronomers who may not be interferometry experts or may not have substantial computing resources at their home institutions.

Compared to other large existing and planned arrays—specifically, the Square Kilometre Array (SKA), in which the United States is not a partner, and ALMA—the ngVLA would provide unique capabilities. These include continuous frequency coverage from 1.2–50 GHz and 70–115 GHz, unique coverage of a key frequency range (15–35 GHz, important for studying terrestrial planet formation, water megamasers associated with $z < 0.5$ galactic nuclei, and cold molecular gas in $z > 2.3$ galaxies), superior point-source sensitivity at all common frequencies compared to the SKA (1.2–15 GHz) and ALMA (35–50 and 70–116 GHz), and access to northern hemisphere sources. In its science reach, the ngVLA would deliver substantial quantitative improvements in the ability to study nonthermal phenomena, a qualitatively new ability to explore the thermal Universe, and a powerful complement to ALMA's higher-frequency capabilities. The ngVLA would realize this science potential by allocating the bulk of its observing time (like ALMA) in response to principal investigator (PI) proposals, rather than (like the SKA) to large, predefined surveys.

M.3.4 Cost, Schedule, and Risks

The ngVLA project team has prepared a detailed project design, plan, schedule, and cost model. Further design and development work is proposed to occupy the next few years, followed by a decade-long construction phase starting in 2025 and a full (steady-state) operations phase running from 2035 through 2054. Costs (including contingency) are estimated by the project team as ≈ $0.1 billion for design and development and ≈ $2.4 billion for construction in 2020 U.S. dollars, translating to ≈ $0.1 billion and ≈ $3.2 billion in then-year dollars. The TRACE analysis estimates design and construction costs summing to ≈ $3.2 billion in 2020 dollars, translating to $4.2 billion in then-year dollars. The RMS panel has arrived at a "hybrid" estimate for the total construction cost that is roughly $130 million (in 2020 dollars) lower than the TRACE value (rounding to ≈ $3.1 billion in 2020 dollars and ≈ $4.2 billion in then-year dollars), but concurs with the TRACE adjustments for (1) a project management and systems engineering "overhead" higher than that for the EVLA upgrade, as appropriate for the ngVLA's larger geographical

extent and greater need for international coordination, (2) schedule threats related to an uncertain time scale for antenna prototyping and a high antenna delivery rate (assumed to be three per month in the steady state) to be sustained over many months, and (3) a higher level of overall contingency (reserve). With regard to (2), the panel notes that the vast majority of the ngVLA's science potential would still be realized if longer integration times were needed to compensate for a modest reduction in the number of antennas—although preserving the concept's excellent *uv* coverage and therefore imaging performance would remain paramount goals. Importantly, all of the above numbers refer to *total* design and construction costs, of which NSF would only contribute 75 percent; given information available as of mid-2020, the panel views prospects for international partner contributions at the desired 25 percent level as excellent. Annual operations costs for the ngVLA are projected to rise from ≈ $147 million in 2035 to ≈ $244 million in 2054 in then-year dollars; again, current expectation is that only 75 percent of these costs would need to be borne by NSF. Adopting the RMS panel's construction cost estimate and a mean annual operations cost of ≈ $100 million in 2020 dollars, the operation-to-construction cost ratio would be ~3 percent, which is at the low end of the envelope filled by previous large projects. The panel appreciates the project team's explicit inclusion of ngVLA decommissioning (at the level of ≈ $0.2 billion in 2020 dollars) in its calculation of total life cycle costs.

Owing to its planned use of mature technology in most areas, the ngVLA would be a project with low technical risk. Technical specifications for the antennas are not overly stringent compared to the current state of the art, or indeed to other observatories that are already operational, although the cost implications of these specifications will become clearer once a satisfactory 18 m antenna prototype exists. Risks related to correlator and receiver development are also low and well understood by the ngVLA project team; risks related to RFI are recognized, and the team is working to develop appropriate mitigation strategies. Last, the RMS panel has considered the risk to the ngVLA's scientific productivity that could arise if its very large image sizes—driven by its large field of view and long baselines—overtax the typical home-institution computing resources of its users. The ngVLA project team understands the scale of this challenge and has plans to address it, for example, via a user-friendly archive coupled with server-side visualization and analysis platforms, although this type of functionality remains an active area of development.

M.3.5 Additional Programmatic Guidance

The RMS panel views the ngVLA as an exciting concept for a flexible, powerful, PI-driven observatory that would address a wide range of Astro2020 high-priority science questions. In support of the project's long-term success, the RMS panel offers three suggestions for its implementation. First, the panel views international participation as essential to the success of the ngVLA project, given the value of sharing technical expertise as well as costs. It would therefore be important for NSF to be proactive in enabling full participation by international partners through both the construction and operations phases of the project. Second, the panel views the growth of a community of future ngVLA users as vital to the ultimate success of the project. In previous decades, growth of the community of future ALMA users was supported by the tight integration of research and training at (sub)millimeter wavelength facilities funded by NSF's University Radio Observatories (URO) program. With the demise of the URO program, the existence of a broad community of ngVLA users tomorrow would require the growth of a broad community of JVLA users today. The RMS panel therefore suggests that concrete progress in making the JVLA accessible to nonexpert users (e.g., via observatory-specified calibration strategies, automatically defined schedules, standard correlator modes, pipeline-reduced data products, and server-side visualization and analysis tools),[3] which can inform detailed design and costing of user interfaces for the ngVLA, factor positively in agency decisions on the start of ngVLA funding. Third, the panel endorses

[3] J. Kern, B. Glendenning, and J. Robnett, 2019, The science ready data products revolution at the NRAO, white paper submitted to the Astro2020 Decadal Survey.

the view that the ngVLA would be a replacement for both the JVLA and the VLBA. Full details of when the JVLA and VLBA would be decommissioned relative to the progress of ngVLA construction remain to be determined, but as a dedicated VLBI array with continent-scale (~10^4 km) baselines, the VLBA is globally unique and will remain crucial for astrometric and other science that demands the monitoring of time-varying phenomena with ultra-high angular resolution. The panel therefore suggests that the VLBA remain operational unless and until its capabilities (ideally, upgraded in the near term by bandwidth increases) are supplied by ngVLA/LBA stations. Continuity here would ensure the existence of a community of future LBA users, and would maintain the VLBI capability that is needed for a number of ongoing long-term (e.g., astrometric) observing programs.

M.4 A LARGE NEW EXPERIMENT: CMB-S4

M.4.1 Introduction

The past decade of measurements of the CMB have yielded precision tests of the ΛCDM paradigm and increasing precision on the parameters that describe the Universe. In the coming decade, sensitive observations of the CMB have the potential to resolve central questions in cosmology, fundamental physics, and particle physics, while also providing new astrophysical insights. The RMS panel supports funding of the CMB-S4 experiment, which is designed to push CMB measurements across critical sensitivity and measurement thresholds to understand the origins of inflation, search for hidden fundamental particles, map out the distribution of mass and hot gas throughout the Universe, and explore time-variable and static millimeter-wave sources. CMB-S4 would apply existing technologies on an unprecedented scale, combining the major ground-based CMB experimental groups and U.S. national laboratories to deliver an instrument matched to scientific need. CMB-S4 is envisioned as a joint NSF and DOE project, has been endorsed by the High Energy Physics community in the 2014 Particle Physics Project Prioritization Panel (P5) report and by the 2015 Academies report *A Strategic Vision for NSF Investments in Antarctic and Southern Ocean Research*, and has achieved DOE “Critical Decision 0,” which confirms the need for investment in this scientific area.

M.4.2 Science Case

The CMB-S4 experiment has been designed by the U.S. and international cosmology communities to address four science themes. The following subsections connect its capabilities to the science questions identified by the Astro2020 science panels.

M.4.2.1 Cosmology

According to the predominant modern theory of cosmology, in the very first moments after the Big Bang, the Universe underwent an exponential expansion known as inflation. Inflation explains key cosmological mysteries, such as the origin of the incredible uniformity of the Universe on very large scales and the measured flatness of space, which would otherwise require a precise tuning of the cosmic energy density. However, the physics that drove inflation has not yet been identified, and understanding what set the Big Bang in motion is now one of the primary cosmology science questions for the coming decade (COS-1). There should be signatures of the inflationary epoch encoded in the CMB that will reveal the origins of this expansion and provide novel information about Grand Unified Theories and quantum gravity, and CMB-S4 is designed to find these signatures.

A clear imprint of the inflation era is a background of gravitational waves echoing through the Universe. These “inflationary gravitational waves” (IGW) introduce vortical patterns (known as “B-

mode," tensor, or curl components) in the observed CMB vector polarization field, which is intrinsically curl-free (and referred to as "E-mode," scalar, or divergence polarization patterns). The ratio of the tensor to scalar polarization modes in the CMB, known as r, encodes the energy scale at which the inflationary expansion occurred (for some classes of models that generate additional B-modes during inflation, the picture is not so simple, but these have particular signatures that allow them to be distinguished). Determining the value of r therefore provides a unique window into the earliest moments of the Universe—temperature/energy scales that are forever beyond collider experiments. Current data indicate that $r < 0.06$, but also prefer families of models that predict $r > 0.001$. Achieving a measurement of r will clarify the underlying physics driving inflation and provide evidence for the quantization of gravity, while limiting r to less than 0.001 at 95 percent confidence will rule out the leading models of inflation. CMB-S4 is designed around achieving this challenging experimental target, matching the target set by the COS science panel (COS-1). It requires a dramatic increase in the number of CMB bolometers in operation, a wide range of independent frequency bands to separate out contaminating foregrounds, and a combination of large and small angular scales to detect the large-scale IGW B-modes and remove contaminating B-modes from lower-redshift lensing of the CMB. The CMB-S4 design is flexible enough to reoptimize its experimental approach during its 7-year lifetime to refine and improve its constraints if a detection of r is made. No CMB experiment less ambitious than CMB-S4 can achieve the needed sensitivity.

The standard cosmological model is a highly successful description of the evolution of the Universe, from the instants after the Big Bang to the present. Within this theory there remain many key details to understand, many of which are tied to fundamental particle physics. The COS science panel has identified the properties of dark matter and the dark sector (COS-2) as a potential breakthrough area for the decade. In particular, the model of a single dark matter particle has given way to a diverse field of potential "Dark Sector" contributions to the energy density of the Universe, expanded sets of particles and fields that are only weakly coupled to known components of the Standard Model of particle physics, predicted as part of extensions to the Standard Model. The CMB provides a unique opportunity to search for the existence of relativistic particles ("dark radiation") that contribute to the cosmic energy density but cannot be sensed in laboratory experiments. For example, light relics imprint measureable perturbations in the acoustic oscillations of the primary CMB temperature and polarization power spectra, while ultralight axions affect the formation of structure on small angular scales, which are detectable in gravitational-lensing induced secondary perturbations to the polarization power spectrum at small angular scales. CMB measurements already demonstrate the reality of the cosmic neutrino background predicted by Big Bang cosmology despite the absence of laboratory detections. To accomplish these goals, CMB-S4 is designed with the angular resolution, sensitivity, and sky coverage needed to precisely measure the perturbations caused by relativistic particles that decouple from the hot early Universe within the first nanosecond (COS-2b), before the quantum chromodynamics phase transition when quarks bind to form hadrons. Such a measurement is not within reach of current experiments, or planned upgrades before CMB-S4.

Understanding the growth of cosmic structure (COS-3) is a third key science question for the decade, and another interface between fundamental particle physics and cosmology. The Universe is suffused with neutrinos, and measurements of neutrino oscillations have demonstrated that these particles have nonzero mass, although these oscillations only determine the differences between the squared masses of the three primary neutrino generations (electron, mu, and tau). The total neutrino mass, and thus the contribution of neutrinos to the energy density of the Universe and their influence on structure formation, remains unknown.

CMB-S4 is designed to achieve a critical threshold (COS-3b) in the measurement of the total neutrino mass—measuring the minimum possible value to 5σ precision. Answering this cosmology science question requires pushing CMB measurements to the limits imposed by cosmic variance, and can only be achieved with an experiment on CMB-S4's scale. CMB-S4 would also make unique measurements of cosmic structure (COS-3a) by mapping the large-scale mass distribution through reconstruction of the CMB lensing potential—enabling fruitful comparisons with tracers of structure at

other wavelengths—and by characterizing the motions of clusters within that structure via the kinetic Sunyaev-Zel'dovich (kSZ) effect.

M.4.2.2 Galaxies, Transients, and the Explosive Universe

While CMB-S4 is designed to deliver precise cosmological measurements, its capabilities open up other science areas that have the potential to engage broader swaths of the astronomical community. CMB-S4 would produce unprecedented maps of ~70 percent of the sky at wavelengths between 1 cm and 1 mm, sampling the full area at least every other day. The sensitivity of these maps would enable a variety of science, particularly the study of hot circumgalactic, intergalactic, and intracluster gas. The temporal sampling would open up this wavelength regime to systematic time-domain studies for the first time.

The scattering of CMB photons by hot electrons results in a characteristic CMB spectral distortion, known as the thermal SZ effect, with an amplitude proportional to the integrated pressure of the hot gas. While this technique has seen its greatest use for the detection and characterization of galaxy clusters, the sensitivity and sky coverage of CMB-S4 maps would make it possible to explore the ionized gas of the circumgalactic medium (GAL-D). Stacking analyses would measure the circumgalactic medium pressure profile to megaparsec radii and constrain the contributions of active galactic nucleus (AGN) and supernova feedback.

As each new wavelength regime (gamma ray, X-ray, optical, infrared, centimeter) has been opened up to systematic time-domain surveys over the past two decades, the number and nature of transient sources have continued to surprise. The millimeter-wave regime probed by CMB experiments is largely unexplored, with only a single, limited experiment over the past decade. Yet the potential sources span a range of exciting possibilities (COEP-2d), ranging from high-redshift and/or orphaned GRB afterglows (peaking in the millimeter regime), to the mysterious fast/blue optical transients like AT2018cow, to tidal disruption events and AGN variability that may be linked to neutrino emission (COEP-D) and black hole accretion physics (COEP-4). Many of these sources peak quickly in the millimeter regime, especially those enshrouded in dust that may be invisible at other wavelengths. Although event rates are uncertain, CMB-S4 would probe all of these with a new combination of cadence and depth, opening up new avenues for follow-up at higher angular resolution with facilities like ALMA and (potentially) the ngVLA, and perhaps also identifying new solar system objects and Galactic transients like stellar flares.

M.4.3 Design Concept

CMB-S4 is designed to take advantage of two well-established millimeter-wave observing sites to conduct two simultaneous surveys. A 7-year ultra-deep survey of 3 percent of sky would take advantage of continuous visibility and outstanding weather conditions at the South Pole. This effort would use 18 Small Aperture Telescopes (SATs), each of diameter 0.5 m, to observe over six bands from 30–270 GHz. In addition, a single Large Aperture Telescope (LAT) of diameter 6m would be used for delensing purposes, operating at 20–270 GHz. In parallel, a 7-year deep/wide survey of 70 percent of the sky would take advantage of the superior sky coverage accessible from the Atacama Desert in Chile. This effort would use two of the same LATs observing at 30–270 GHz. All ~500,000 detectors required would be of the transition edge sensor type, with cryogenic multiplexing readouts.

M.4.4 Cost, Schedule, and Risks

The CMB-S4 project team has prepared a detailed project design, plan, schedule, and cost model, which are compatible with NSF and DOE protocols for management of large projects and have already

been refined through several rounds of internal review. First light is proposed for 2026, with the end of construction in 2028 leading to a full (steady-state) operations phase that concludes in 2035. Rigorous costing by the team in accordance with DOE methodology implies costs (including contingency) of ≈ $30 million for design and development and ≈ $500 million for construction in 2020 U.S. dollars, summing to ≈ $600 million in then-year dollars. The TRACE analysis estimates design, development, and construction costs summing to ≈ $660 million in 2020 dollars, translating to $700 million in then-year dollars. The RMS panel has arrived at a "hybrid" estimate for design, development, and construction costs that is slightly higher than the project team's (≈ $560 million in 2020 dollars), concurring with the TRACE adjustments for (1) a higher assumed "overhead" for information technology, computing, and software during construction, and (2) schedule threats related to the timely fabrication of an unprecedented number of cryogenic detectors. With regard to (2), the panel notes that impacts on schedule and cost are substantially reduced by the 1 year of schedule contingency that is already built into the CMB-S4 project plan. In the construction phase, DOE:NSF cost sharing is expected to be in the ratio 7:5, implying that the panel's estimated total cost for design, development, and construction would translate to costs of ≈ $330 million in 2020 dollars (≈ $370 million in then-year dollars) to DOE, and ≈ $230 million in 2020 dollars (≈ $260 million in then-year dollars) to NSF. These estimates make no assumptions about cost savings that might be possible if CMB experiments aligned with CMB-S4 were to make in-kind contributions of infrastructure.

A preliminary bottom-up estimate by the CMB-S4 team implies an annual operations cost of ≈ $33 million in 2020 dollars (≈ $55 million in then-year dollars) after averaging over the experiment's nominal 7-year lifetime. In this phase, DOE:NSF cost sharing is tentatively expected to be in the ratio 1:1, with further sharing of the NSF portion among multiple divisions under consideration. Adopting the RMS panel's construction cost estimate and a mean annual operations cost of $100 million in 2020 dollars, the operation-to-construction cost ratio would be ~6 percent. The CMB-S4 project plan notes that normal end-of-life decommissioning costs for South Pole and Chile infrastructure are anticipated.

Owing to significant heritage from previous generations of CMB experiments, including the ongoing third-generation Simons Observatory (SO) and South Pole Observatory (SPO), CMB-S4 would be a project with medium/low technical risk. The primary source of programmatic risk is the challenge of scaling up to a high production rate across multiple fabrication sites in order to deliver a large number of cryogenic detectors on a tight timeline. As noted above, a year of contingency in the project schedule already provides substantial mitigation on this front. Through a dedicated working group guided by external reviews, the CMB-S4 project team is exploring other risk reduction strategies—for example, enlisting more facilities beyond the planned three DOE labs in the detector fabrication effort. The RMS panel has concluded that the scaling challenge here is not insignificant, but that prospects for mitigation are good.

M.4.5 Additional Programmatic Guidance

The RMS panel views CMB-S4 as a powerful, cosmology-focused experiment that would address Astro2020 priority science questions at a level that no other concepts can. In support of the project's long-term success, the RMS panel offers the following two suggestions for its implementation. First, the panel suggests that third-generation CMB experiments aligned with CMB-S4—specifically, the SPO and the "nominal" version of the SO—be high priorities for federal support.[4] Besides training students and postdoctoral researchers, thereby empowering them to play vital future roles in CMB-S4, these

[4] The RMS panel considered whether intermediate sensitivities (superior to the third-generation CMB experiments' and inferior to CMB-S4's) would be worth pursuing as a separate decadal goal. However, facilities delivering such sensitivities would not by themselves address the Astro2020 COS panel questions, and would find it difficult (in terms of schedule) both to build on lessons from third-generation experiments and to inform CMB-S4 strategy.

experiments are poised to help retire technical risk for CMB-S4 and usefully inform its strategies for surveying the sky and removing foreground signals. Second, the panel views it as appropriate for an experiment at the cost scale of CMB-S4 to be more "observatory-like" in seeking broad engagement with astronomers beyond the traditional CMB community, and ensuring that (for example) plans for data management and event alerts maximize opportunities for transient science to the extent possible without sacrificing the primary cosmology goals. The panel therefore suggests that an articulated plan for engaging the broader astronomical community be a precondition for the start of CMB-S4 funding.

M.5 SIGNIFICANT FUNDING TO SUPPORT MID-SCALE PROJECTS

Over the past decade, funding of projects at the mid-scale level (for NSF, currently defined as costing $2 million to 70 million, with awards made via competitive proposal calls) has become an important, agile, and cost-effective mechanism for enabling the construction and operation of world-class RMS facilities in a variety of science areas, while training the next generation of scientists and instrument builders. Funded mid-scale projects have included the EHT, which can image the environs of supermassive black holes in and beyond the Milky Way, and multiple experiments that explore the early Universe by probing the CMB or the Epoch of Reionization (EoR), when the intergalactic medium transitioned from being mostly neutral to mostly ionized. Funded projects have also included new instruments or improvements for large single-dish telescopes operating in observatory mode, including a more accurate surface for the GBT, a new multi-pixel camera for Arecibo, and a new multi-band millimeter camera for the Large Millimeter Telescope (LMT). Most of these awards have been made through the Mid-Scale Instrumentation Program (MSIP), which was established by the Division of Astronomical Sciences in response to a recommendation of the Astro2010 decadal survey. Very recently, two awards in support of RMS projects have been made through the broader Mid-scale Research Infrastructure-1 (MSRI-1) program, which was established in support of one of NSF's ten "Big Ideas." The substantial oversubscription of both MSIP and MSRI-1 funding is an indicator of high demand in this cost range and high quality of funded projects. That RMS projects in particular have competed so successfully in these programs reflects the wealth of scientific opportunities in this wavelength regime.

Based on the ambition and creativity of the Astro2020 white papers, the RMS panel is confident that funding for mid-scale projects will be as valuable and as impactful in the next decade as it has been in the past. In particular, based on the Astro2020 high-priority science questions, the RMS panel has identified four areas in which outstanding scientific opportunities exist for new mid-scale RMS facilities. These areas are discussed in detail below, with reference to the specific white papers that inspired them, although since each still requires navigating a complex path to a successfully competed MSRI-2 (up to $70 million) proposal, the panel is highlighting them as exciting opportunities rather than endorsing concepts exactly as presented. In three of the four areas, future investment would build on previous MSIP and/or MSRI-1 funding. The panel views the limited previous investment in the fourth area (solar broadband imaging) as a missed opportunity, given that earlier versions of the Frequency Agile Solar Radiotelescope (FASR) concept were strongly endorsed by the Astronomy and Astrophysics decadal surveys in 2000 and 2010 and the Solar and Space Physics decadal surveys in 2002 and 2012, but only the subset of its capabilities represented by the Expanded Owens Valley Solar Array (EOVSA) have been implemented. It would be important to structure future mid-scale funding competitions so that research in areas like ground-based solar physics is not inherently disadvantaged by the fact that it is pursued across more than one NSF division.

In order of nearest to most distant observational target(s), a first key area of mid-scale opportunity is broadband, spectropolarimetric imaging of the Sun. To date, solar radio observations from EOVSA and the Murchison Widefield Array (MWA) have made significant advances using spectro-imaging observations of the Sun, revealing the evolving spatial and energy distributions of high-energy electrons in flares, mapping spatial and temporal changes of the coronal magnetic field, tracing the origin of coronal heating, and performing 3D mapping of the magnetic field in sunspots. However, each facility has

limitations: because of its small number of antennas, EOVSA does not have the dynamic range needed for high-fidelity imaging of rapidly time-varying phenomena or the low-frequency coverage needed to study coronal plasma emission, while the MWA is not a dedicated solar array and cannot provide the observing time or infrastructure needed for a comprehensive view of the Sun. Bastian et al. (2019)[5] present a concept for FASR, a facility optimized (in terms of bandwidth, sampling rate, angular resolution, and imaging dynamic range) to study the extreme ranges of flux density and temporal variations of the Sun over a frequency range from 0.2–20 GHz. Such a dedicated solar facility would allow daily imaging of the dynamic solar atmosphere from the middle chromosphere through the solar corona with a cadence of several times per second, and would have the ability to image narrower frequency bands with time resolution as fine as 20 ms. To achieve this performance, two separate arrays of antennas would be spread over ~3 km footprints (not necessarily at the same site): a ~64 element array of 2m antennas operating from 2–20 GHz, and a ~48 element array of 6m antennas operating at 0.2–2 GHz. Such a facility would be spectacularly powerful for understanding the dynamic atmosphere of the Sun, solar activity (SSSP-1), and all key components of space weather (SSSP-4), sampling both thermal plasma and nonthermal particles and uniquely sensitive to solar magnetic fields (SSSP-3). It would also be an invaluable partner to solar space-based missions in the coming decade. The arrays would launch a new era of "4 D" studies of the Sun through dynamic imaging spectroscopy with unprecedented spatial, temporal, and frequency resolution (SSSP-D) to probe the evolution of complex solar phenomena and the couplings between them. Such spectro-imaging of the Sun at radio wavelengths would reveal the extraordinarily complex range of phenomena that occur over various spatial and temporal scales within stellar atmospheres and directly probe how these processes affect the physics and dynamics of the solar atmosphere, including the temperature structure, the circulation of material, the driving of winds, and the ejection of plasma through coronal mass ejections (CMEs). These insights in turn would inform our understanding of the behavior of other stars (SSSP-3). While the basic concept remains well-aligned with Astro2020 priorities, significant changes in technology have occurred since FASR was first proposed. A logical first step toward a full MSRI-2 scale proposal would be a redesign of the original 2010 concept (via MSIP/MSRI-1 funding) that takes into account technological advances and cost-savings opportunities driven by commercial developments over the past decade.

A second key area of mid-scale opportunity is high-resolution imaging of jets driven by supermassive black holes in the centers of galaxies. The current state of the art in terms of resolution is provided by the EHT, an experiment that has regularly combined a number of telescopes around the world (including ALMA) for VLBI observations at a relatively short (1.3 mm) wavelength. Because angular resolution is proportional to observing wavelength and inversely proportional to separations between telescopes, the EHT has been able to deliver an unprecedentedly sharp view of the center of the galaxy M87 (see Figure M.1c). Doeleman et al. (2019)[6] present a concept for an expansion of the EHT that would entail (a) building ten new 10 m diameter telescopes at additional sites around the world, in order to improve image fidelity, and (b) quadrupling the recording bandwidth, in order to improve sensitivity and enable simultaneous polarimetric imaging at 1.3 mm and 0.87 mm. By virtue of its improved imaging performance and higher angular resolution, such a facility would provide essential insights on the question of how jets are formed and powered (COEP-3). M87, whose supermassive black hole drives a powerful jet at 99 percent of the speed of light and has already been imaged by the EHT at an angular resolution comparable to its projected Schwarzschild radius, would be a uniquely promising target for a more capable facility. By making sensitive, multi-frequency, spatially and temporally resolved observations with higher dynamic range, it would be possible to shed light on the details of how magnetic

[5] T. Bastian, H. Bain, R. Bradley, B. Chen, J. Dahlin, E. DeLuca, J. Drake, et al., 2019, Frequency agile solar radiotelescope: A next generation radio telescope for solar astrophysics and space weather, white paper submitted to the Astro2020 Decadal Survey.

[6] S. Doeleman, L. Blackburn, J. Dexter, J.L. Gomez, M.D. Johnson, D.C. Palumbo, J. Weintroub, et al., 2019, Studying black holes on horizon scales with VLBI ground arrays, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1909.01411.

fields extract rotational energy from the black hole and/or its surrounding accretion disk to drive M87's jet. Larger samples (of tens to hundreds of systems) would enable studies of jet physics on scales farther from the Schwarzschild radius, and through the EHT's ability to resolve binary black holes at high angular resolution would also help address the question of how supermassive black holes grow (COEP-4, GAL-3). A logical first step toward a full MSRI-2 scale proposal here would be completion of the design and prototyping work that has recently been funded through an MSRI-1 award to the EHT team.

A third key area of mid-scale opportunity is the surveys of the static and time-variable radio sky that would be enabled by an innovative new "radio camera" instrument. Technological advances in low-noise, room-temperature amplifiers and commercial computer power and networking have made it possible to conceive of large (of order ~1000 element) arrays of radio dishes whose signals are combined to produce science-ready images in real time without deconvolution—that is, a true radio camera. An array operating in the GHz range with baselines extending out to 15 km would have angular resolution of a few arcseconds. Hallinan et al. (2019)[7] present the Deep Synoptic Array 2000 (DSA-2000) concept, consisting of 2000 × 5 m steerable dishes covering the entire 0.7–2 GHz frequency range and designed from the ground up for survey science. Such a radio camera could efficiently survey the entire observable sky with multiple pointings over multiple epochs, produce full-Stokes maps with noise well below a μJy/beam, and reveal of order a billion radio sources. The resulting data sets would address a large and diverse subset of the Astro2020 high-priority science questions. An enormous catalog of FRBs, triggered and communicated in real-time, would enable exploration of the diversity of explosive phenomena across the electromagnetic spectrum (COEP-2). Time-domain searches would also contribute to searches for radio afterglows of compact object mergers detected by LIGO and Virgo (COEP-D), for CMEs from other stars (SSSP-4), and potentially for technosignatures as tracers of life on exoplanets (EAS-D). A radio survey camera with the large collecting area of the DSA-2000 concept could make a very significant contribution to pulsar timing efforts in support of gravitational wave detection (see below). Beyond time-domain science, a broadband radio survey of the entire sky to unprecedented depth would enable studies of how supermassive black holes form and grow in concert with their host galaxies (GAL-3), and searches for the signatures of dark matter annihilation (COS-2). A logical first step toward a full MSRI-2 scale proposal here would be the further development of a design (building on the DSA-2000 team's existing prototype array, recent MSIP funding for a 110-element precursor, and ongoing characterization of possible sites in the American West) and supporting partnerships that would be compatible with the upper limit on mid-scale funding by NSF.

A final key area of mid-scale opportunity, probing the largest cosmic scales, is mapping the evolution of HI in the early Universe. Parsons et al. (2019)[8] discuss the current status of this exciting field, in which a detection of the earliest phase of the transition from a mostly neutral to mostly ionized intergalactic medium has recently been reported by the Experiment to Detect the Global EoR Signature (EDGES). Several other current experiments, including the Hydrogen Epoch of Reionization Array (HERA), Phase II of the MWA (MWA-II), and the Large-Aperture Experiment to Detect the Dark Ages (LEDA) are in the process of obtaining data that could deliver the first HI power spectrum of the EoR and potentially confirm the EDGES result interferometrically. While this field is in its infancy, the potential impact of opening an entirely new window on cosmic evolution is enormous, and there is considerable value in having independent experiments with different designs (and therefore different observational systematics). Technical lessons and scientific results from these projects, supplemented by further prototyping of hardware, software, and analytical techniques (e.g., refined methods for in-situ mitigation of potential systematic errors, new array elements, direct imaging FFT correlators, and real-time calibration), would inform the design of one or more next-generation experiments by the end of the

[7] G. Hallinan, V. Ravi, S. Weinraub, J. Kocz, Y. Huang, D.P. Woody, J. Lang, et al., 2019, The DSA-2000: A radio survey camera, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1907.07648.

[8] A. Parsons, J.E. Aguirre, A.P. Beardsley, G. Bernardi, J.D. Bowman, P. Bull, C.L. Carilli, et al., 2019, A roadmap for astrophysics and cosmology with high-redshift 21 cm intensity mapping, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1907.06440.

decade. Possible architectures could include an "EoR imager" focused on cross-correlations with multi-wavelength probes of structure at redshifts $z < 12$, and a "Cosmic Dawn Array" focused on $z > 12$ power spectrum measurements where non-HI probes are unavailable. Connections to high-priority science questions are clear: characterization of HI in the early Universe would directly constrain the thermal history of the intergalactic medium and the topology of reionization (GAL-1), and open up the use of the pre-reionization "Dark Ages" as a cosmological probe (COS-D). Constraints on the 21cm power spectrum would also aid understanding of the distribution of dark matter on small scales (COS-2). A logical first step toward a full MSRI-2 scale proposal here would be assimilation of the results and lessons from HERA and MWA-II en route to the design of one or more next-generation experiments. The RMS panel notes that insights on observational systematics gained from the current generation of EoR experiments—for example, related to beam characterization and coupling between neighboring antennas—would also inform the design of possible future HI intensity mapping experiments targeting lower redshifts, such as the Packed Ultra-wideband Mapping Array (PUMA) that was presented to Astro2020 in conceptual form.

The above set of mid-scale opportunities reflects not only the diversity of RMS science that can be supported at this level of investment, but also the diversity of phases in which different projects may find themselves relative to the decadal survey cycle. By issuing calls for mid-scale proposals on a regular basis, NSF would accommodate projects that become funding-ready at different points in the decade.

M.6 CROSSCUTTING CAPABILITIES: PULSAR TIMING, INSTRUMENTATION DEVELOPMENT, AND RFI MITIGATION

M.6.1 Pulsar Timing, and Continuing Support for Arecibo and the Green Bank Telescope

Pulsars are highly magnetized, rapidly rotating neutron stars whose beams of emission represent some of the most stable clocks in the Universe. By making exact measurements of the arrival times of pulses from individual pulsars and large networks of pulsars, astronomers can draw conclusions about the properties of the neutron stars and the spacetime through which the pulses travel on the way to Earth. To be successful, pulsar timing programs must satisfy several important criteria: (1) they must involve observations at *multiple frequencies*, so that the effects of intervening interstellar gas on pulse arrival times can be corrected; (2) they must be *long-term*, to enable more precise measurements of secular changes in those arrival times (owing to the effects of precession, general relativity, or low-frequency gravitational waves); (3) they must be observed with an *uninterrupted cadence*, to prevent the loss of information on phasing and/or long-wavelength gravitational wave sensitivity; and (4) in the case of pulsar timing networks, they must be accompanied by pulsar *search* programs that can identify new objects for inclusion in those networks in order to improve sensitivity to gravitational waves. With the exception of (4), where large single-dish telescopes or very closely packed arrays are needed for efficient searches, these criteria can in principle be satisfied by many possible combinations of current (Arecibo, GBT, JVLA) and potential future (ngVLA, mid-scale radio survey camera) RMS facilities.

From the Astro2020 science panel reports, it is clear that pulsar timing capabilities are critical for tackling a number of high-priority science questions: the mass and spin distributions for neutron stars and black holes (COEP-1), the growth of supermassive black holes (COEP-4, GAL-3), the synthesis of information from electromagnetic, particle, and gravitational wave signals (COEP-D), the properties of dark matter (COS-2; pulsar timing arrays can potentially detect anomalies owing to lensing by dark matter lumps), and the cosmological implications of gravitational waves (COS-4). From the RMS panel's interaction with the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) team, which is a global leader in the use of pulsars for gravitational wave detection, it is equally clear that large amounts of observing time with sensitive facilities will be critical for increasing the size of the timing sample from its current ~80 to 200 millisecond pulsars by the end of the decade. (A "minimum observing program" for NANOGrav would require ~1200 hours a year of timing observations with broadband receivers on facilities that include at least one with Arecibo-scale collecting area, plus ~1000 hours a year

for pulsar searches; an "ideal observing program" would use up to 2.5× as many hours for timing at a higher cadence.) Given the uncertainties in whether and when the ngVLA or a mid-scale radio survey camera would become available to contribute to pulsar timing efforts, the desire to support continuing U.S. leadership in this area, the need to avoid any gaps in timing coverage, and the fact that Arecibo and the GBT are the only existing or proposed U.S. facilities that can effectively contribute to pulsar *search* efforts, the RMS panel views continued operational support for Arecibo and the GBT over the next decade as essential.

While pulsar timing represents their clearest science driver in the context of Astro2020, both Arecibo and the GBT are also poised to deliver abundantly in addressing other high-priority science questions. Used as auxiliary VLBI stations, both facilities greatly improve the sensitivity of high-resolution observations of jets launched by neutron star mergers, tidal disruptions, and other explosive events (COEP-2, COEP-3). Arecibo's planetary radar capability plays a vital role in characterizing small solar system bodies, whose properties inform the understanding of debris disks around other stars (EAS-1, EAS-3). Sensitive single-dish observations complement the capabilities of the JVLA (and potentially the ngVLA), allowing detections of faint emission lines from sources too extended or too low in surface brightness to be studied by interferometric arrays (ISM-1, ISM-2). To deliver on their broad scientific potential, it is important for both observatories to have stable bases of NSF (and, if available, state) funding that can support healthy fractions of peer-reviewed "open time" scientific observations.[9] Time purchases by outside partners will continue to be important going forward, but it is clear from recent experience that large commitments to such partners can deprive observatories of scheduling flexibility and jeopardize their ability to execute high-priority fixed-time observations of the sort described above (e.g., for VLBI, radar, and the timing of pulsars outside the NANOGrav network).

M.6.2 Instrumentation Development

As discussed above, the past decade's scientific advances at RMS wavelengths have been strongly enabled by technological advances and the work of talented instrument builders (including software developers) to leverage them. Maintaining the capacity for instrumentation development across the U.S. astronomical community is essential for the future health and progress of the field. The challenge is to align funding with projects that (1) address technology needs, (2) are on a scale where students and postdoctoral researchers can engage, and (3) have platforms for deployment. While MSIP and MSRI funding of mid-scale projects plays a valuable role here, smaller-scale projects and cutting-edge technology development efforts are also important, and are well matched to dedicated funding via smaller grants (e.g., as provided by the NSF Advanced Technologies and Instrumentation program, which has supported EDGES and new instruments for the GBT and LMT, among other projects). Even if funded, however, new instrumentation can be exploited only if there are facilities where it can be deployed. Historically, this role has been filled mainly by single-dish telescopes that can accommodate a range of guest and/or facility instruments. Recent examples include Arecibo, the GBT, and the LMT, although the closure of the Caltech Submillimeter Observatory (CSO) means there is now a U.S. "deployment capability gap" at the shortest RMS wavelengths. User-contributed instrumentation is also increasingly compatible with large aperture synthesis arrays, with digitized voltage streams from individual telescopes packaged into standard Ethernet packets and broadcast to multiple backends for different commensal uses (radio transients, pulsars, technosignatures, etc.). For both single-dish and array observatories, the RMS panel suggests that ability and willingness to accommodate the deployment of user-contributed instrumentation (encompassing both hardware and software) factor positively in discussions of federal funding levels. The panel also notes that the scientific potential of a new instrument can only be fully exploited if all of its software needs—including, when relevant, the processing of extremely large data

[9] For the GBT, whose wide range of high-priority observing modes creates an unusual level of scheduling complexity, a "healthy fraction" would be at least 50 to 60 percent.

sets—receive robust long-term support. Overall, continued support for the development and deployment of new instruments, the development and continued support of requisite data processing software, and the training of the next generation of instrument builders, is a key RMS investment for the next decade and beyond.

M.6.3 Mitigation of Radio Frequency Interference

A significant challenge for all ground-based facilities operating at radio wavelengths in the coming decade will be the need to contend with the growing problem of RFI from human-made sources. The rapid increases in the quantity, bandwidth, and power of RFI from terrestrial sources and satellite constellations pose an existential threat to radio astronomy.[10] Without suitable mitigation efforts, RFI will increasingly impact the ability to detect spectral line emission from atoms and molecules at all redshifts, as well as faint sources of thermal and nonthermal continuum emission that require averaging over large frequency bandwidths.

Protection of the radio sky requires a multi-faceted approach, and there are a number of ways that funding agencies can support this effort. One is to advocate for protection of radio observatories from sources of RFI geographically, spatially, and temporally. Examples of geographical and spectral protection include the preservation of existing radio quiet zones and protected (passive) frequency bands, respectively. Temporal separation includes the exploration of coordinated dynamical sharing of the spectrum between various users. In addition, the RMS panel encourages agencies to provide adequate funding to all current and future RMS facilities for the development of RFI protection and mitigation strategies, including specially designed hardware and sophisticated software tools to excise RFI without indiscriminately deleting signals from real astronomical sources (e.g., temporally varying fast radio bursts and other transients). The panel also encourages agencies to increase the levels of funding they already provide to astronomers and scientific institutions for training, advocacy, and public communication on RFI threats and mitigation approaches. Among these efforts are continued advocacy for scientific use of the spectrum as part of the overall management of the radio spectrum as a shared resource through the work of the National Academies Committee on Radio Frequencies (CORF) and other U.S. and international organizations.

M.7 GUIDING PRINCIPLES

From the earliest stages of the Astro2020 process, multiple stakeholders within the astronomical community have encouraged the development of an ambitious, exciting, and scientifically motivated program for the next decade. The RMS panel has taken this encouragement to heart in arriving at the set of investments described above, which it is offering for the consideration of the Astro2020 steering committee. To inform the steering committee's deliberations on which selections to order from the lengthy "menu" defined by the program panels' reports, the RMS panel also offers three top-level principles governing its overall vision. First, the panel views it as important that facility operations budgets for the next decade include full support for the U.S. share of ALMA. As discussed above (and reflected in Table M.1), ALMA is a productive and scientifically vibrant observatory, which has already engaged an impressively broad swath of the global astronomical community and is poised to make further progress on many of the next decade's high-priority science questions. Continuing operations funding over the next decade would enable further facility improvements within the envelope of the current development budget, with more ambitious and costly improvements possible in future decades. Second, the panel expects that science will flourish best with a program of investments extending over large,

[10] L. van Zee, D. DeBoer, D. Emerson, T.E. Gergely, N. Kassim, A.J. Lovell, J.M. Moran, et al., 2019, Spectrum management, white paper submitted to the Astro2020 Decadal Survey.

medium, and small cost scales, properly balanced so that bigger investments do not crowd out their smaller cousins. This principle is informed by the history of discovery in astronomy, which shows that major disruptive discoveries are made in diverse ways—by individuals and small groups, or by large teams; using modest observing resources, or operating on the cutting edge; driven by serendipity and human inspiration, or achieved through dogged persistence. More generally, literature citation patterns demonstrate that discovery and development in science are strong functions of team size: large teams tend to excel in developing existing ideas, whereas small, agile, risk-tolerant teams are more likely to make disruptive discoveries.[11] The panel's support of large new facilities is therefore intertwined with its support for traditional individual investigator grants, opportunities for small teams to pursue ambitious observing programs, funding to support the development of new technologies and instrumentation, and facilities that offer opportunities for student training and/or substantial amounts of open time for diverse, risk-tolerant investigations. These more modest investments offer the potential for outsized science return.

The panel's third governing principle relates to the constructive, respectful, and substantive engagement of the astronomy community with stakeholders from outside that community in addressing environmental and cultural concerns, including at intersections with indigenous rights. The design and construction of new ground-based facilities operating at RMS wavelengths offer opportunities to set high standards for professional astronomers' interactions with indigenous communities, allowing consent to emerge from a sustained and genuinely collaborative process. A rigorous environmental impact assessment can provide an initial sense of the full spectrum of concerns about a new facility, but for all RMS facilities, ongoing consultation with community stakeholders is necessary to minimize negative impacts of operations on surrounding areas and associated cultural activities. To make sure that telescope sites are ultimately returned to their original conditions, projects need to understand and budget for all decommissioning activities before receiving construction funding. To help ensure that RMS facilities' impacts are as positive for their immediate communities as for society at large, the panel suggests that agencies provide funding for meaningful stakeholder engagement at all phases of the project life cycle.

[11] L. Wu, D. Wang, and J.A. Evans, 2019, Large teams develop and small teams disrupt science and technology, *Nature* 566:378–382.

N

Report of the Panel on the State of the Profession and Societal Impacts

N.1 SYNOPSIS

Six astronomy and astrophysics decadal surveys have been produced by the National Academies of Sciences, Engineering, and Medicine to date. While each discussed the state of the profession and societal impacts, Astro2020 is the first to feature a formal panel devoted to these concerns. The statement of task given to the panel was as follows:

> The Panel on State of the Profession and Societal Impacts will gather information on the health and demographics of the astronomy and astrophysics community and make actionable suggestions to the Astro2020 committee on the topics of demographics, diversity and inclusion, workplace climate, workforce development, education, public outreach, and relevant areas of astronomy and public policy. The panel's suggestions will be incorporated into a program for all of astronomy and astrophysics by the Astro2020 committee.

This is a new era in astronomy and astrophysics on every measurable axis, with many indicators of substantial progress in the past decade. Major new observatories and space-based telescopes are poised to produce massive quantities of new data about the cosmos. Near daily coverage of astronomical discoveries in the popular media—images of black holes, discoveries of potentially habitable worlds—reveals the field's effective communication with the public. The number of students pursuing degrees in physics and astronomy continues to grow, and the field is becoming more representative of U.S. demographics, with steady increases in the number of women and Hispanic Americans.

A deeper look, however, reveals a Profession (Box N.1) with profound weaknesses. At many undergraduate programs, the attrition rate of physics and astronomy students is high. Nationally, students from underrepresented groups interested in physical sciences are less likely to complete physics and astronomy majors than white students, leading to the persistent underrepresentation of community members in the field relative to their representation in society at large. Racial discrimination and sexual harassment continue, leading to a climate in the field that depresses recruitment of women, people of color, and people from other traditionally minoritized groups, and an increase in the proportion of people from those groups who leave the field at all levels.

BOX N.1 The Profession

The community of scientists, engineers, technicians, and nontechnical people engaged in the production and instruction of astronomical knowledge, as well as learners on the path to joining their ranks.

Women and people of color, people with disabilities, LGBTQIA+, nonbinary people, and people who hold two or more of these identities remain extremely underrepresented in senior leadership positions. Furthermore, astronomers have not always engaged adequately with local communities

impacted by observatories; the consequences are made evident by the growing resistance from Indigenous peoples and their supporters, particularly surrounding construction on the summit of Maunakea (Mauna Kea) in Hawai'i, which is considered sacred by Indigenous people. This perspective has prioritized facilities over their impact on people and cultures and is facing increased resistance from those most impacted. The health of the Profession depends on setting priorities and doing science in ways that are mindful of these human realities. The promise of the decadal survey is the collective power of the Profession to carry out an ambitious, actionable plan to assert its priorities and, where needed, to change its trajectory. However, previous decadal surveys did not include written insight into nontechnical factors that guided their ranked selection: the human-centered processes necessary to carry out the science. The panel asserts that fundamentally, the pursuit of science, and scientific excellence, is inseparable from the humans who animate it. This statement guides the panel's suggestions.

The panel proposes that by explicitly pursuing a set of equity-advancing values, in addition to articulating priorities for scientific investment, the Profession will improve the quality of science overall in the next decade and beyond. Equitable access, multimodal expertise, responsible stewardship, and accountability are four values that are defined and discussed in Section N.5. These values are reflected in the founding documents of the federal agencies that support astronomy and are reflected in best practices reported in the literature, white papers and public town halls for the decadal survey, the expertise of the panel, and numerous National Academies reports and consensus studies. For the Profession to maximize scientific advancement over the coming decade, the panel suggests that resources must be invested toward realizing these equity-advancing values by fostering engagement, increasing opportunities for participation by the full human diversity of our nation, and laying the foundation for lasting change. These values embody the panel's vision for the Profession by 2030. This report provides a clear articulation of these values, along with suggestions for implementation and assessment. The panel identifies *seven essential goals* for the Profession in the next decade:

1. *Collecting, Evaluating, and Acting on Demographic Data:* Collect and report consistent demographic data from organizations that support astronomical research, education, and training. Data are key to identifying promising practices, measuring progress, and holding agencies and institutions accountable to equity-advancing values.
2. *Leveraging Power:* Use funding structures to recognize and realize equity-advancing values.
3. *Reimagining Leadership:* Develop, select, and sustain diverse cohorts of leaders who lead by exercising equity-advancing values.
4. *Addressing Harassment and Discrimination:* Establish clear policies, collect and report relevant metrics, and enforce accountability measures to remove structures and individuals that perpetrate identity-based discrimination (including harassment) in astronomy.
5. *Removing Barriers:* Modernize practices that have a disparate impact on access to education, training, and advancement.
6. *Cultivating Local and Global Partnerships:* Reframe policies around community engagement in order to embed cultural humility, ethical practice, and a growth mindset throughout the Profession in a continuous effort to cultivate and sustain healthy cultures for scientific inquiry.
7. *Partnering with Indigenous Communities:* Align the values of the Profession with those of Indigenous and other local communities impacted by the Profession to cultivate and sustain healthy partnerships for the benefit of both.

The panel suggests methods for funding agencies, professional societies, university departments, observatories, research institutes, government laboratories, and the Profession to embody these values and achieve these goals. The panel estimates[1] that these methods range from no-cost to fairly substantial

[1] The programmatic suggestions included in this report have been derived by the panel on a best-effort basis that includes examination of the costs of existing similar programs, consultation with agency staff, and other relevant

investments and provides agencies with specific guidance. *On the order of $40 million per year spread across the National Aeronautics and Space Administration (NASA), National Science Foundation (NSF), and Department of Energy (DOE) would be required to address the highest priorities*. Ideally, this funding would not come at the expense of current research grant funding but would be supplementary. Not all issues can be addressed immediately. Where appropriate, the panel identified methods to be implemented rapidly and others that require more time.

As this report was written in mid-2020, the United States was in the midst of profound self-examination of social and economic inequalities resulting from historic and systemic racism, discriminatory police brutality highlighted by the Black Lives Matter movement,[2] sexual harassment and inequalities highlighted by the #MeToo movement, and the starkly inequitable and severe health and economic impacts of the COVID-19 pandemic on people of color. This background of social ferment and introspection makes all the more timely and more urgent a frank assessment of the ties between the equity and well-being of the Profession, including issues of race, ethnicity, gender, and workplace climate.

N.2 THE LANDSCAPE OF THE ASTRONOMY PROFESSION

This section gives a snapshot of the state of the Astronomy Profession as of mid-2020, the forces that have shaped it over the past decade, and those likely to shape it going forward. This discussion is not comprehensive, but outlines themes and conditions of the Profession that ground the panel's suggestions.

N.2.1 What Does the Panel Mean by "the Profession"?

Historically, astronomers have conceived the Profession (Box N.1) as a relatively dehumanized scientific enterprise, pursuing observatories and data with secondary regard to the humans who use them and the values that animate their work. As a result, progress toward an equitable and inclusive profession that is representative of the population has been slow. Astronomers have often failed to ethically engage with communities who are impacted by the facilities that they build. Barriers to equitable access and advancement are ingrained in both educational and professional astronomy contexts. Identity-based discrimination and sexual harassment continue within the Profession. These wrongs undermine the professional integrity and the scientific excellence of the Profession.

The panel asserts that people and organizations are integral to the discovery process. The panel is concerned with the full scope of resources that enables scientific advancement. This includes the Profession that *does* the science, their knowledge and skills (i.e., human capital), and equity in the organizations where the work is carried out. In this report, the panel considers "the Profession" to be the community of scientists, engineers, technicians, and nontechnical people engaged in the production, dissemination, and instruction of astronomical knowledge, as well as learners on the path to joining their ranks. More than half (54 percent) of full-time employed U.S. American Astronomical Society (AAS) members with Ph.D.s work at institutions of higher education; 33 percent work at government labs, research institutes, or observatories.[3] Also relevant are the broader set of communities with whom the Profession interacts, including amateur astronomers, editors, journalists, and educators who enable, support, communicate, and inspire the work of astronomy.

Astronomy is a quest to understand the universe and humanity's place within it. Its discoveries resonate deeply with the public. Many in the Profession depend on a small number of shared resources:

---

documentation. The specific resources used in making individual suggestions are provided in footnotes. The final implementation of suggestions made here will necessarily reflect more formal and thorough analysis of cost, schedule and most effective programmatic practices.

[2] See Box N.2, "Black Lives Matter."

[3] J. Pold and R. Ivie, 2019, "Workforce Survey of 2018 US AAS Members Summary of Results," https://aas.org/sites/default/files/2019-10/AAS-Members-Workforce-Survey-final.pdf, accessed 26 August 2020.

observatories and supporting infrastructure. While most astronomy funding comes from NASA and NSF, astronomy is one of many priorities for these agencies.[4] Still larger societal forces profoundly shape the Profession. Yet, as a small field, astronomy can be nimble and experimental, and grass-roots efforts of a few individuals through policy change can have a greater impact on the whole of the Profession. For example, NASA rapidly switched guest observer programs to dual-anonymous proposal review to reduce implicit bias from the review process.[5] Owing to astronomy's outsized visibility and influence in public opinion, its move toward equitable and inclusive practices may influence other professions to move in the same direction.

With these core qualities of the Profession in mind, the next section considers the central role of investments and the impacts they have on the Profession. Then, the panel provides a summary of the Profession's demographics at different points in the pathway from college to and through career and in both academic institutions and research labs.

N.2.2 Investments and Their Impacts

Funding affects everything from technology and infrastructure to academic opportunities and human capital development. Therefore, funding decisions influence which members participate, advance, and feel they belong. Overall funding for astronomy in the past decade has been increasing, with growing investment in facilities. But support of astronomers performing research and funding to train future researchers has been flat or declining during this same period.

N.2.2.1 Federal Agencies

Funding from federal agencies for astronomy has grown 40 percent in the past decade; however, this falls far short of the doubling in federal investment in astronomy that was envisioned in the 2010 decadal survey. Furthermore, funding for individual investigators and proposal success rates have been flat or declining. The Astronomy and Astrophysics Advisory Committee (AAAC) reports why this is happening and the impact on science and scientists of the declining proposal success rates.[6] The NSF Division of Astronomical Sciences (AST) allocates only 17.5 percent of its total division budget to its primary individual investigator grants program, the Astronomy and Astrophysics Research Grants (AAG). The largest share of the division's budget has historically been directed toward facility operation. Meanwhile, the grant success rate for NSF AST fell from 50 percent in 1990 to close to 15 percent in 2015 and has remained under 20 percent since. NSF AST undertook a mid-decade review to identify opportunities for divestments. This led to the elimination of NSF's Partnerships in Astronomy and Astrophysics Research (PAARE), their sole program to develop human capital at undergraduate and graduate levels through partnerships with minority-serving institutions. NASA Science Mission Directorate (SMD) proposal funding rates were a healthy 30–35 percent in the early 2000s, but since then they have steadily declined.

[4] For example, in the fiscal year 2020 NASA budget, the $1.729 billion spent on astrophysics was 7.6 percent of the total budget, and 24 percent of the NASA Science budget. The NSF Division of Astronomical Sciences budget was $287 million out of the NSF's $8.578 billion budget. DOE's Cosmic Frontier budget was $94.9 million out of DOE's $38.586 billion budget.

[5] A. Witze, 2019, NASA changes how it divvies up telescope time to reduce gender bias, *Nature* 571:156.

[6] P. Cushman, et al., 2015, "Impact of Declining Proposal Success Rates on Scientific Productivity," https://arxiv.org/ftp/arxiv/papers/1510/1510.01647.pdf.

N.2.2.2 Private Foundations

Private philanthropy has been an important source of funding for astronomy for over a century.[7] For instance, the Carnegie Institute funded the development of the Mt. Wilson Observatory in 1904, and construction of the Hale Telescope on Palomar Mountain was funded by the Rockefeller Foundation in 19282The Sloan Digital Sky Survey, supported by the Sloan Foundation, has transformed how astronomical research is conducted. The Heising-Simons Foundation recently funded the PI Launchpad to increase the number of space mission proposals led by principal investigators (PIs) with historically marginalized identities. A growing number of private foundations and individual philanthropists fund ground-based optical telescopes, individual researchers through fellowships and award programs, university participation in observatories, and an increasing role supporting postdoctoral researchers.[8]

N.2.2.3 Impacts of Trends in Investment

Scarcity of funding threatens capacities for creativity and risk-taking that are essential to bold scientific advancement. It also negatively affects the culture of workplaces and the training of the next generation. Scarcity increases the likelihood that both everyday and scientific decisions will be driven less by an ethical vision of scientific conduct than by urgency and pressure.

In this environment, the panel sees at least three promising directions that would enable progress. First, the field can communicate funding priorities to federal agencies to increase direct support to astronomers as researchers, mentors, and communicators, relative to agencies' funding of facilities. Second, private foundation support could grow significantly to play a bigger role in the Profession's future. Of the $2.3 billion in private funds distributed to science in 2017, 87 percent went to life sciences, with 11 percent ($250 million) going to physical sciences.[9] Third, the Profession can associate more closely with industry and related fields with strong growth and investment, such as data science and advanced computation, to better support a variety of career paths.[10] There are many applied areas where astronomy is positioned to contribute to the training of the scientific workforce.

N.3 DEMOGRAPHIC LANDSCAPE

The past decade has witnessed a substantial growth in the desire of Americans to participate in the excitement of astronomical discovery. The number of astronomy B.S. and Ph.D. degrees shows continued growth (Figure N.1). As nearly daily coverage of astronomical discoveries in the popular media reveals, the field is effectively communicating with the public. While there has been a steady increase in the numbers of women and Hispanic American degree recipients (Figures N.1 and N.2), the number of African American students earning Ph.D. degrees remains low and unchanged over three decades (Figure N.2).

[7] This section was informed by data collected by the Astro2020 Panel on an Enabling Foundation for Research.
[8] See https://www.hsfoundation.org/programs/science/51-pegasi-b-fellowship/.
[9] This information was provided by Marc Kastner, president, Science Philanthropy Alliance, in a presentation to the Enabling Foundation for Research Panel, 22 October 2019.
[10] The need for academic institutions to do a better job informing students about alternative career options was one of the recommendations of the Astro2010 decadal report.

American Institute of Physics (AIP) statistics show the unemployment rate of 2014–2016 astronomy Ph.D.s to be only 3 percent,[11] similar to other STEM fields.[12] Those joining the private sector with a bachelor's degree or Ph.D. earn a median starting income of $60,000 and $120,000, respectively, which is higher than most other fields.[13] A significant driver of these employment outcomes is the increasing importance of computational skills and data-science approaches in astronomy training and research. Newly minted Ph.D.s in astronomy are now more likely than ever to forgo a postdoctoral appointment and enter the nonacademic research workforce.[14] The Profession has great capacity for enabling an array of excellent career outcomes in defense, healthcare, or commerce, as well as teaching.

At the same time, broader demographic trends reveal a systemic failure of the Profession to attract, retain, and advance diverse talent. About 2.5 percent of all first-year white students compared to about 1.5 percent of African American/Black, Hispanic/Latino and American Indian/Alaska Native[15] first-year students intend to major in the physical sciences.[16,17] While 11 percent of white students intending to major in the physical sciences will earn a degree in physics and astronomy, only 4 percent of students from underrepresented groups with similar intent complete physics and astronomy degrees.[18] This is consistent with earlier findings that only 40 percent of students who enter university with an interest in science, technology, engineering, and mathematics (STEM) and 20 percent of STEM-interested underrepresented minority students finish with a STEM degree.[19] Engagement of American Indian/Alaska Native people in astronomy at the undergraduate level is the lowest of all physical sciences, with an average of 2 individuals receiving bachelor's degrees per year.[20] Since astronomical first light on Maunakea 50 years ago, there have been a total of three Ph.D.s in astronomy or astrophysics awarded to Native Hawaiians. The loss of Black, Hispanic, and Indigenous physics and astronomy students during their undergraduate years is reflected in the low percent entering graduate school and the fewer than 10 astronomy Ph.D.s (out of nearly 200 Ph.D.s; see Figures N.1 and N.2) produced annually. This failure of the *undergraduate* educational system has long-term consequences for the diversity of the Profession at the doctoral level and beyond.

The 2015 and 2019 Inclusive Astronomy Conferences have raised attention on the experiences of LGBTQIA+ astronomers and astronomers with disabilities. In 2018, 1 percent of AAS members

---

[11] American Institute of Physics, "Physics Trends: Astronomy Ph.D.s One Year Later," https://www.aip.org/T/physics-trends/astronomy-phds, accessed 19 May 2021.

[12] National Science Foundation, National Center for Science and Engineering Statistics, "Survey of Doctorate Recipients," 2015, https://ncsesdata.nsf.gov/doctoratework/2015/html/SDR2015_DST_4_1.html, accessed 19 May 2021.

[13] P. Mulvey and J. Pold, "Astronomy Degree Recipients One Year After Degree," https://www.aip.org/statistics/reports/astronomy-degree-recipients-one-year-after-degree, accessed 26 August 2020. Comparisons to other fields can be found at https://www.aip.org/statistics/physics-trends/what-do-new-bachelors-earn for bachelor's degrees and at https://ncsesdata.nsf.gov/doctoratework/2015/html/SDR2015_DST_53.html for Ph.D.s, accessed 19 May 2021.

[14] P. Mulvey and J. Pold, "Astronomy Degree Recipients One Year After Degree," https://www.aip.org/statistics/reports/astronomy-degree-recipients-one-year-after-degree, accessed 26 August 2020.

[15] The terms used in the paragraph are not the choice of the panel. Rather, they are a consequence of the terms used in federal data collection, on which this analysis draws.

[16] Unfortunately, the number of entering first-year students who intend to major in physics/astronomy is not known.

[17] Appendix Table 2-16, "Freshmen Intending S&E Major by Field, Sex, and Race or Ethnicity, 1998–2012," NSF Science and Engineering Indicators, 2016, https://www.nsf.gov/statistics/2016/nsb20161/#/report/chapter-2/undergraduate-education-enrollment-and-degrees-in-the-united-states, accessed 21 May 2021.

[18] Calculated from AIP Enrollments and Degrees Survey, various years, unpublished.

[19] PCAST (President's Council of Advisors on Science and Technology), 2012, "Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics," https://eric.ed.gov/?id=ED541511.

[20] L. Merner and J. Tyler, 2017, "Native American Participation Among Bachelors in Physical Sciences and Engineering," AIP focus on March 2017.

identified as a gender other than man or woman, and a few astronomers identified as transgender. In the same 2018 survey of 1,915 AAS members, 85 percent identified as heterosexual or straight, while 7 percent of AAS members identified as gay, lesbian, or bisexual, leaving 7 percent of the membership identifying as other (2 percent) or preferred not to respond (5 percent). In a 2016 survey of AAS members, 94.7 percent of the membership responded as not having any of the disabilities listed (hearing, sight, or mobility issues), with 2.5 percent preferring not to respond. In 2018, a more expanded survey of disabilities among U.S. AAS members was prepared.[21] Recognizing additional disabilities, as well as providing the open-ended option for individuals to select “other disability,” reduced the number to 82 percent of AAS members not identifying with a disability.

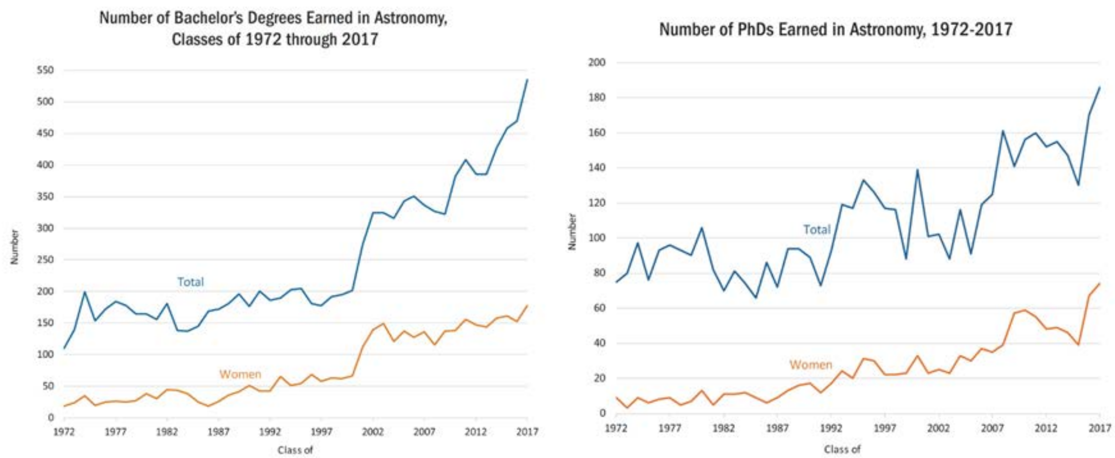


FIGURE N.1 *Left:* Number of bachelor’s degrees, total and to women, earned in astronomy, 1972–2017. *Right:* Number of Ph.D.s, total and to women, earned in astronomy, 1972–2017. SOURCE: Porter and Ivie (2019).

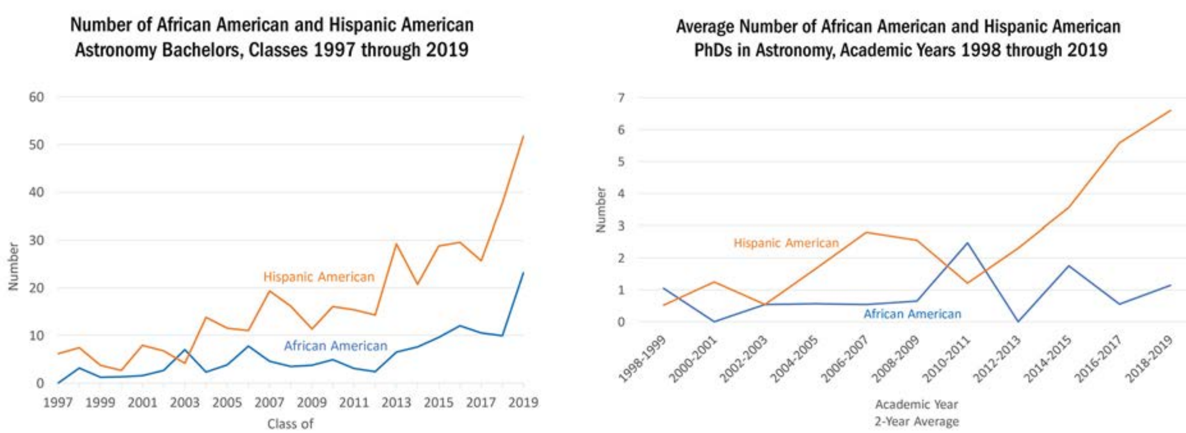


FIGURE N.2 The numbers of astronomy degrees earned by African American and Hispanic American students. *Left:* Bachelor’s degrees. *Right:* Doctorates. SOURCE:
Courtesy of the Statistical Research Center at the American Institute of
Physics.

[21] J. Pold and R. Ivie, 2019, “Workforce Survey of 2018 US AAS Members Summary of Results,” https://aas.org/sites/default/files/2019-10/AAS-Members-Workforce-Survey-final.pdf, accessed 26 August 2020.

N.3.1 Academic Institutions

The current landscape for racial and ethnic diversity among astronomy faculty remains dismal. African American people comprise a mere 1 percent of the faculty over all ranks among astronomy departments; Hispanic people comprise just 3 percent.[22] Their collective representation of 4 percent is eight times below these groups' joint representation in the U.S. population.[23] The low representation of Black undergraduate physics and astronomy students[24] indicates that no significant increase at higher academic levels can be expected for more than a decade. This underrepresentation was identified as a problem as far back as the 1980 decadal survey.[25] Representation of these groups is slightly better in physics departments, although they are not uniformly distributed among the nation's colleges and universities. Indeed, as of 2016 there was only one astronomy department that had representation of both African American and Hispanic American faculty, and roughly two-thirds of astronomy departments had representation of neither.[26]

Gender representation among astronomy faculty has improved over the past decade. In 2014, women in assistant and associate professor ranks each comprised 29 percent, up from 23 percent in 2003. Women make up 15 percent of full professors, up from 10 percent in 2003.[27] Recent AIP surveys show that the low representation is owing not only to the "lag" effect of fewer Ph.D.s to women in the past. Systematic gender differences in family work and career-advancing opportunities and resources have disadvantaged women in physics and astronomy.[28]

N.3.2 Large Research Facilities

Large research facilities, managed through organizations (AURA, AUI, USRA[29]) in cooperative agreements with the funding agencies, complement colleges and universities as significant employers of the professional astronomical workforce. They hold real potential to improve the Profession by aligning their composition with society and in training the future workforce. The organizations contacted by the panel expressed a positive stance toward diversity and inclusion; however, they could not identify specific mechanisms for monitoring or holding accountable the aspects of their operations bearing on diversity, equity, or inclusion. Both AURA and AUI have diversity officers and documented engagement with the broader community.

Because NASA manages its centers internally, it can directly implement change. The NASA Science Plan (2020)[30] states, "As research has shown, diversity is a key driver of innovation and more diverse organizations are more innovative. … NASA believes in the importance of diverse and inclusive teams to tackle strategic problems and maximize scientific return." The panel identified positive steps

---

[22] AIP Academic Workforce Survey, 2016, unpublished results.

[23] The 2019 Census gives 18.5 percent of the U.S. population as Hispanic or Latino and 14 percent Black or African American, for a joint representation in the United States of 32.5 percent.

[24] The AIP National Task Force to Elevate African American Representation in Physics and Astronomy (TEAM-UP), 2019, *The Time Is Now: Systemic Changes to Increase African Americans with Bachelor's Degrees in Physics and Astronomy*, College Park, MD: American Institute of Physics.

[25] Field, et al., *1980 Decadal Survey*, vol. 1, Appendix B, p. 172, and vol. 2, starting on p. 334.

[26] AIP Academic Workforce Survey, 2016, unpublished results.

[27] J. Pold and R. Ivie, "Workforce Survey of 2018 US AAS Members Summary of Results," https://aas.org/sites/default/files/2019-10/AAS-Members-Workforce-Survey-final.pdf, accessed 26 August 2020.

[28] A.M. Porter and R. Ivie, 2019, "Women in Physics and Astronomy," https://www.aip.org/statistics/reports/women-physics-and-astronomy-2019, accessed 26 August 2020.

[29] AURA: Association of Universities for Research in Astronomy; AUI: Associated Universities, Inc.; USRA: Universities Space Research Association.

[30] NASA Science Mission Directorate, 2020, "Explore Science 2020–2024 A Vision for Science Excellence," https://science.nasa.gov/science-pink/s3fs-public/atoms/files/2020-2024_Science-TAGGED.pdf.

NASA has taken to this end, such as instituting dual-anonymous proposal review to reduce risk of bias. However, additional targeted procedures and accountability mechanisms are needed, as well as transparency of the demographic representation of its staff and individuals they serve.

N.4 THE EVOLVING LANDSCAPE

In the past two decades, the field has undergone massive shifts in the structure and size of research teams, the places where research is carried out, and the skill sets for which students are trained. Large collaborations and survey-scale missions are increasingly prominent, with an explosion of data and a workforce that is more digitally connected and more geographically distributed than ever before. The "grand challenges" of the next 10 years will require advanced, innovative methodological and computational approaches to solve. It is imperative that the current and coming generations of astronomers are trained in computational methods. Despite the broad access to massive data sets via public facilities and surveys, the most powerful computers, and the knowledge and training to use them, is not openly accessible. Institutions where most astronomers and students from underrepresented groups reside have the least access and thus least opportunity to engage in this new mode of astronomical training and discovery.

The Profession also needs to modernize its core structures, such as its approaches to mentoring (from master-apprentice models to evidence-based practices such as mentoring networks), research funding (from structures that disadvantage underresourced institutions to a more equitably distributed funding model that rewards individuals who can carry out needed systemic changes), leadership (from a system that overburdens individuals from underrepresented groups to one that elevates them), education (from traditional lecture to inclusive pedagogy and research-based instruction), and community engagement (from unidirectionally broader impacts to mutually beneficial community partnerships).

Meanwhile, the demographics of the United States are changing; communities of color already constitute the majority in states such as California. Efforts to accelerate the participation of racially minoritized populations in astronomy graduate education such as the Cal-Bridge Program and the Fisk-Vanderbilt Master's-to-Ph.D. Bridge Program are engaging America's population in the Profession. These programs are two examples of ways in which astronomy has become a leader in STEM by implementing structural changes toward equity, diversity, and inclusion. The AAS Task Force on Diversity and Inclusion in Graduate Education made recommendations based on lessons learned from research and from these and other initiatives to improve graduate admissions, recruitment, and mentoring, as well as program climate and data use.[31] Particularly amid disruptions to both testing and test preparation owing to COVID-19, many graduate programs have eliminated both general and physics GRE score requirements for Ph.D. admissions in order to more effectively attract talented, high-achieving students from an increasingly diverse pool of candidates.[32] The emerging sensibility around equity-based holistic review has applicability not only for admissions, but also hiring, awards, grants, and leadership positions.

Women in Astronomy (WIA) have organized (most recently in 2017) meetings to equip one another and the Profession with knowledge, perspective, research, and skills to advocate for equitable workplaces. Concerns about focus on gender without consideration of intersections with race and other identities led to the 2015 and 2019 Inclusive Astronomy conferences, which gathered astronomers, social scientists, and policy makers to highlight issues of equity and inclusion from an explicitly intersectional perspective. The AAS endorsed the recommendations following these meetings. Recently, the AIP commissioned a report from the National Task Force to Elevate African American Representation in

[31] Recommendations from the Task Force can be found at https://aas.org/sites/default/files/2019-09/aas_diversity_inclusion_tf_final_report_baas.pdf.

[32] See https://www.sciencemag.org/careers/2020/06/graduate-programs-drop-gre-after-online-version-raises-concerns-about-fairness.

Undergraduate Physics and Astronomy (TEAM-UP).[33] Based on student and department head surveys, site visits to high-performing physics departments, and interviews with African American students, the report identified five key factors for their success: (1) fostering a sense of belonging, (2) creation of a physics identity, (3) effective teaching and academic support, (4) personal financial support, and (5) leadership to create environments, policies, and structures to maximize African American student success. The report's goal is to at least double the number of African American physics and astronomy bachelor's degree recipients by 2030, and to transform the field into one that welcomes, includes, and values all people.[34]

The panel believes that the Profession must apply the same level of regard it demonstrates in studying the universe toward the recognition of how human and social interactions contribute to knowledge and workplace culture. Just as the Profession relies on its technicians to constantly update computational resources and software to keep pace with science capabilities, it must also work with social scientists to help it foster the social, psychological, structural, and cultural environment where all who ponder the universe can share in the production and validation of that knowledge.

N.5 A VALUES STATEMENT FOR THE PROFESSION OF ASTRONOMY AND ASTROPHYSICS

Guiding Principle: The pursuit of science, and scientific excellence, is inseparable from the humans who animate it.

The construction of the astronomy and astrophysics decadal survey is itself a demonstration of the inseparability of humans and the science they produce. Each decadal survey, delivered for congressional review, is the product of an internal set of weighted priorities determined by a selected group of people with an assumed broad set of astrophysical expertise. This interplay between (1) the *people* who influence and create the Profession; (2) the *processes*, norms, structures, and systems that govern and are reflected by everyday activities; and (3) the consensus-based *priorities* and desired outcomes they advocate for, are artifacts of the culture[35] of astronomy. All scientific disciplines are knowledge communities, with ways of knowing, distinctive language, and beliefs about what types of questions are most important to pursue. The knowledge that communities produce, including scientific communities, are therefore cultural knowledge.[36] Stated another way, "scientific knowledge is but a particular form of cultural knowledge."[37]

As recent National Academies reports[38] have emphasized, *the climates and cultures of science are inseparable from the advancement of scientific knowledge and investments*. Previous decadal surveys

---

[33] See https://www.aip.org/sites/default/files/aipcorp/files/teamup-full-report.pdf.

[34] The AIP National Task Force to Elevate African American Representation in Physics and Astronomy (TEAM-UP), 2019, *The Time Is Now: Systemic Changes to Increase African Americans with Bachelor's Degrees in Physics and Astronomy*, College Park, MD: American Institute of Physics.

[35] Culture is the "languages, customs, beliefs, rules, arts, knowledge, and collective identities and memories developed by members of all social groups that make their social environments meaningful" (American Sociological Association). See https://www.asanet.org/topics/culture#:~:text=jpg,make%20their%20social%20environments%20meaningful, accessed 18 August 2020.

[36] R.K. Merton, 1973, quoting Max Weber, in "Science and the Social Order (1938)," in *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago: University of Chicago Press.

[37] L. Chambers, 2019, A different kind of dark energy: Evidence for placing race and gender in physics, white paper submitted to the Astro2020 Decadal Survey, https://baas.aas.org/pub/2020n7i162/release/1.

[38] NASEM (National Academies of Sciences, Engineering and Medicine), 2018, *Sexual Harassment of Women*, Washington, DC: The National Academies Press; NASEM, 2019, *The Science of Effective Mentorship in STEMM*, Washington, DC: The National Academies Press; NASEM, 2018, *Graduate STEM Education for the 21st Century*, Washington, DC: The National Academies Press; NASEM, 2020, *Addressing the Underrepresentation of Women in Science, Engineering, and Medicine*, Washington, DC: The National Academies Press.

have prioritized technical resources to produce the best science, but they did not include written insight into the nontechnical factors that guided their ranked selection. This resource prioritization includes the implicit values that have historically guided astronomy's culture and are demonstrated by its composition, behaviors, and rituals. While implicit cultural values can be learned over time, this favors those already living within these norms and thus embeds structural inequity. By leaving implicit the values that have guided past decadal surveys' scientific priorities, the Profession failed to move forward together toward a more equitable future, as described Section N.2, above. Therefore, an explicit statement of equity-advancing values—(1) equitable access, (2) multimodal expertise, (3) responsible stewardship, and (4) accountability, which are further described below—will enable the Profession to foster engagement, increase opportunities for equitable participation in the field, and lay the foundation for lasting scientific excellence in a more diverse nation.[39,40,41,42]

For example, the Profession's inherently hierarchical structure, based on assumed individual superiority of innate scientific capacity, perpetuates in part by casting the structure of opportunity as a "scientific meritocracy." Meritocracies are well-known to reproduce structural inequities by defining merit using metrics that favor historically privileged groups and disadvantage those with different or emerging forms of leadership and expertise.[43] The Profession demonstrates commitment to scientific rigor in its pursuit of understanding the universe by conceptualizing and launching successful missions, as prioritized in this and previous decadal surveys. However, the Profession has not prioritized equitable access to the resources available from federal sponsoring agencies in pursuit of that understanding, as evidenced by the large gap between the demographic profile of the Profession and the U.S. population.[44] Given that intellect is distributed equally across the entire human population, any deviation from the country's demographic composition in the Profession delineates structural inequities[45] along the pathways to full participation in the field. These structural inequities include, but are not limited to, racism, sexism, ableism, homophobia, xenophobia, neurotypical bias, and the intersecting oppression, often with multiplicative deterring effects on those holding multiple marginalized identities. The history of this country and the Profession encode discrimination and structural oppression into virtually every system.[46,47] The Profession must proactively remove these systems and replace them with evidence-based, equity-advancing ones. The field must evolve narrow definitions of scientific rigor into scientific excellence. The panel defines scientific excellence as the equitable optimization of knowledge, infrastructure, and innovations, and includes technical and nontechnical contributors and stakeholders,

---

[39] S.A. Hewlett, M. Marshall, and L. Sherbin, 2013, How women drive innovation and growth, *Harvard Business Review–HBR Blog Network*, August 23.

[40] A. Kezar, 2013, *How Colleges Change: Understanding, Leading, and Enacting Change*, New York: Routledge.

[41] A.J. Kezar and E.M. Holcombe, 2017, *Shared Leadership in Higher Education*, Washington, DC: American Council on Education.

[42] A. Kezar and D. Maxey, 2014, Faculty matter: So why doesn't everyone think so, *Thought and Action* 2014:29–44.

[43] J.R. Posselt, 2016, *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*, Cambridge, MA: Harvard University Press.

[44] Population Census, 2019, U.S. demographics (White: 60 percent; Asian or Asian American: 6 percent; Hispanic or Latino: 18.5 percent; Black or African American: 14 percent; American Indian or Alaska Native: 1.5 percent), https://www.census.gov/quickfacts/fact/table/US/PST045219.

[45] C. Miller and K. Stassun, 2014, A test that fails, *Nature* 510:303–304.

[46] C. Miller and K. Stassun, 2014, A test that fails, *Nature*, 510:303–304.

[47] S.C. Wilder, 2013, *Ebony and Ivy: Race, Slavery, and the Troubled History of America's Universities*, New York: Bloomsbury Press.

which produce higher quality[48,49,50,51] and more innovative[52] outcomes. By this definition, true scientific excellence is not possible without equitable participation.[53]

There is no better time to take stock of the Profession than in a drastically changing world. The impacts of the "COVID era"[54] have not been experienced in recent history: millions of deaths worldwide, widespread shelter-in-place, and a severe recession. Add to this the international demonstrations of millions of people in support of the Movement for Black Lives—a response to often unprosecuted police-sanctioned and vigilante murders of Black people (see Box N.2). Taken together, the sense that the world is irreversibly changing cannot be denied, and as members of the astronomical and global community, the Profession is unquestionably affected by these events. In this context, the relationships between product and person and between individual and community are being interrogated in unprecedented ways that are directly relevant to the survival and success of professional astrophysics.

Clearly articulated values are necessary to guide policy development. This is demonstrated by the founding and amended legislation (hereafter, the founding documents) that outline the congressionally mandated goals of the sponsoring agencies. The sponsoring agency values[55] identified are: (1) innovation (e.g., NASA Act 2008; Section 102.d.5[56]); (2) economic prosperity (e.g., NSF Act 2018; Section 1862.a.1[57]); (3) health and well-being (e.g., DOE Act 2014; Section 7111.2[58]); and (4) broadening participation (e.g., DOE Act 2014; Section 7141.b[59]). These values connect the well-being of individuals in and adjacent to the Profession to the national interest.

---

[48] H.H. Friedman, L.W. Friedman, and C. Leverton, 2016, Increase diversity to boost creativity and enhance problem solving, *Psychosociological Issues in Human Resource Management*, 4(2):7–33.

[49] K.A. Jehn, G.B. Northcraft, and M.A. Neale, 1999, Why differences make a difference: A field study of diversity, conflict and performance in workgroups, *Administrative Science Quarterly*, 44(4):741–763.

[50] T.H. Cox and S. Blake, 1991, Managing cultural diversity: Implications for organizational competitiveness, *The Executive*, 5(3):45–56, *JSTOR*, www.jstor.org/stable/4165021, accessed 18 August 2020.

[51] S.A. Hewlett, M. Marshall, and L. Sherbin, 2013, How women drive innovation and growth, *Harvard Business Review–HBR Blog Network*, August 23.

[52] B. Hofstra, B., et al., 2020, The diversity–innovation paradox in science, *Proceedings of the National Academy of Sciences* 117(17):9284–9291.

[53] B. Hofstra, B., et al., 2020, The diversity–innovation paradox in science, *Proceedings of the National Academy of Sciences* 117(17):9284–9291.

[54] E. Yong, 2020, How will the coronavirus end? *Atlantic*, 25 March, https://www.theatlantic.com/health/archive/2020/03/how-will-coronavirus-end/608719/, accessed 18 August 2020.

[55] The sponsoring agency values identified here are not an exhaustive list of the priorities as outlined in their respective founding documents, but are indicative of priority convergence as determined by this panel.

[56] NASA Act, 2008, "The preservation of the role of the United States as a leader in aeronautical and space science and technology and in the application thereof to the conduct of peaceful activities within and outside the atmosphere."

[57] NSF Act, 2018, "The Congress finds that the fundamental research and related education program supported by the Federal Government and conducted by the Nation's universities and colleges are essential to our national security, and to our health, economic welfare, and general well-being."

[58] DOE Act, 2014, "The Congress of the United States finds that this energy shortage and our increasing dependence on foreign energy supplies present a serious threat to the national security of the United States and to the health, safety and welfare of its citizens."

[59] DOE Act, 2014, "The Director shall have the duty and responsibility to advise the Secretary on the effect of energy policies, regulations, and other actions of the Department and its components on minorities and minority business enterprises and on ways to insure that minorities are afforded an opportunity to participate fully in the energy programs of the Department."

BOX N.2 Black Lives Matter

The members of the State of the Profession and Societal Impacts Panel unequivocally affirm that Black lives matter. This needs to be manifested in the treatment of Black people indirectly and directly connected to this Profession. Current data[1] and the historical record[2] tell us that much progress remains urgently to be made. The brutality and systemic racism that have resulted in an ever-growing list of slain Black people, including[3] George Floyd, Ahmaud Arbery, and Breonna Taylor, have necessitated mass demonstrations and strikes in which many individual astronomers and astronomical organizations and societies have participated.

To those of us who identify as both Black and astronomers, the traumatizing effects of anti-Black racism are no different whether experienced inside or outside the classroom, in the laboratory or on the street getting to and from the places where we do our science.[4] We cannot bring our full and best selves to astronomical inquiry when simply getting to work can be frightening, injurious, or fatal. Many existing reports,[5] first-person accounts,[6,7] and news coverage[8] make it clear that the same anti-Black racism that exists in society also persists within professional spaces and professions—including astronomy.[9.10]

This is but one facet of a complex and overlapping web of the institutionalized marginalization and oppression of many people influencing and influenced by the Profession.[11] All aspects of structural oppression need to be thoughtfully and consistently addressed before we can truly say that we have an inclusive field of astronomy and astrophysics.

---

[1] See https://mappingpoliceviolence.org/aboutthedata, accessed 26 August 2020.

[2] See https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6080222/, accessed 26 August 2020.

[3] See https://interactive.aljazeera.com/aje/2020/know-their-names/index.html, accessed 26 August 2020.

[4] E. Armah, 2012, Emotional justice, *Network Journal*, 19(2):68.

[5] See https://www.nature.com/articles/d41586-020-01741-7, accessed 26 August 2020.

[6] See https://twitter.com/BlackInTheIvory.

[7] See https://www.nature.com/articles/d41586-020-01741-7, accessed 26 August 2020.

[8] See https://www.chronicle.com/article/I-Was-Fed-Up-How/248955, accessed 26 August 2020.

[9] See https://www.buzzfeednews.com/article/stephaniemlee/university-florida-astronomy-racism-emails, accessed 26 August 2020.

[10] See https://www.buzzfeednews.com/article/stephaniemlee/yale-astronomy-systemic-racism-emails, accessed 26 August 2020.

[11] See https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017JE005256?_ga=2.68586385.1242262924.1598481211-1505064369.1598481208, accessed 26 August 2020.

Motivated by these documents, the panel explicitly defines a set of human-centered, equity-advancing values that present an opportunity for current and potential astronomers to equitably contribute to scientific excellence. These equity-advancing values are: (1) equitable access, (2) multimodal expertise, (3) responsible stewardship, and (4) accountability, and the panel describes them here with their connection to sponsoring agency values. These values were not developed in isolation, but were gleaned from best practices as reported in the literature (described below), from submitted white papers, public town halls for the present decadal survey, expertise of the panel, and numerous National Academies reports and consensus studies. In short, the panel has sought to summarize, synthesize, and clarify a minimum set of values that would increase equity in the field of astronomy and astrophysics.

Some readers may find the terms and concepts included in this values section new and uncomfortable. However, extensive sociological, psychological, pedagogical, and organizational development literature indicates that it is precisely this confrontation of new concepts and ideas that leads to unexpected insights. The panel notes that there is a strong evidentiary basis in research on change about the importance of sensemaking, that is, making new meaning around concepts through a variety of

inputs.[60,61] The reader is urged to attempt to inhabit a different and perhaps unfamiliar perspective in trying to interpret these ideas. These short paragraphs represent decades of research, experiences, and expertise, only a small fraction of which can be presented here. The reader is invited to explore the literature cited here and throughout the report.

Equitable access is the set of practices, procedures, norms, and structures that ensure that all current and future astronomy community members can contribute their unique talents and perspectives to the field, while having fair use of all necessary and available resources. An equitable professional structure has to include an "analysis of existing decision-making, agenda-setting power structures and the degree to which those structures are proximal to those with collective, implementation-level knowledge of how those large-scale decisions impact the members with lowest institutional power."[62] This requires the identification and acknowledgement of the ways that individuals and institutions contribute to scientific excellence and how relationships among various actors impact the Profession. Equitable access accounts for the extent to which there is equity *within the organizations* that educate the next generation, for whom training is a prerequisite to access career opportunities in the Profession. Cultivating equitable access allows the Profession to fully contribute to the agencies' value of equitable participation by anyone in the United States.

Multimodal expertise is the multiple ways of prioritizing, assessing[63] and evaluating knowledge, including the science and research objectives of the field. In the humanities, it is termed "epistemology."[64] Multimodal learning[65,66] and leadership[67] styles are well known in the social sciences, STEM education, and management/leadership literatures. Such expertise will expand the scope of inquiry in unexpected ways,[68]and require the development of broader skill sets for many members of the Profession. Valuing it will also require recognition of many who already demonstrate such skills and leadership abilities. This includes technical, individual, interpersonal, cultural, and systems-thinking practices such as active listening, open-mindedness, attention to universal design,[69] cultural humility[70] and literacy, social justice, and growth mindsets. The Profession historically prizes objective, rational thinking in the pursuit of scientific rigor. Yet, how it approaches research questions, determines funding priorities, locates sites of research and investigation, conducts the research, and interprets the results are all dependent on the

---

[60] See https://www.aip.org/sites/default/files/aipcorp/files/teamup-full-report.pdf

[61] S. Elrod and A. Kezar, 2016, *Increasing Student Success in STEM*, Washington, DC: Association of American Colleges and Universities.

[62] M. Dones, et al., "Systems Transformation," *National Innovation Service*, https://www.nis.us/systems-transformation, accessed 18 August 2020.

[63] L. Winig and R. Livingston, "Values-Based Leadership Across Difference: The Life and Legacy of Nelson Mandela," *Harvard Business Review*, https://values-based-leadership-across-difference-the-life-and-legacy-of-nelson-mandela/KS1238, accessed November 9, 2020.

[64] S.G. Harding, ed., 1987, *Feminism and Methodology: Social Science Issues*, Bloomington: Indiana University Press.

[65] E.F. Keller, 1984, *A Feeling for the Organism, 10th Anniversary Edition: The Life and Work of Barbara McClintock*, New York: Macmillan.

[66] A. Lightman, 2018, *Searching for Stars on an Island in Maine*, New York: Vintage.

[67] J. Alvehus, 2019, Emergent, distributed, and orchestrated: Understanding leadership through frame analysis, *Leadership*, 15(5):535–554, doi:10.1177/1742715018773832.

[68] J. Posselt, 2016, *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*, Cambridge, MA: Harvard University Press.

[69] "Universal design is the design and composition of an environment so that it can be accessed, understood, and used to the greatest extent possible by all people regardless of their age, size, ability, or disability." See https://universaldesign.ie/Built-Environment/Shared-Space/Shared-Space-Full-Report.pdf.

[70] Cultural humility is a "lifelong commitment to self-evaluation and self-critique, to redressing the power imbalances … , and to developing mutually beneficial and non-paternalistic clinical and advocacy partnerships with communities on behalf of individuals and defined populations." M. Tervalon and J. Murray-Garcia, 1998, Cultural humility versus cultural competence: A critical distinction in defining physician training outcomes in multicultural education, *Journal of Health Care for the Poor and Underserved*, 9(2):117–124.

communities of individuals that are animating those processes. This becomes increasingly important as large collaborations become more prominent. Increasing the number of perspectives, expertise, experiences, and cultural touchpoints makes the process of collaborative work more difficult,[71] but the outcomes more just,[72,73] innovative,[74] and of higher quality.[75,76,77,78] This productive friction will encourage scientific excellence while increasing equitable access. Multimodal expertise maps directly to the sponsoring agency value of innovation. The Profession's increasing complexity will require broader skill sets than simply technical expertise; valuing these skills will produce innovative outcomes.

Responsible stewardship is the reciprocal care for the environment, land, and people in relation to resources consumed by the Profession. Responsible stewardship requires that the Profession recognize the "*common but differentiated responsibilities and respective capabilities*"[79] held throughout astronomy. The actions taken by the Profession impact living beings and the environment. Practicing responsible stewardship in the Profession can lead to the agencies' value of economic prosperity by supporting the learning and development of its membership and prioritizing environmentally, financially, and socially responsible scientific inquiries.

Accountability is the clear articulation of thoughtful, rigorous, site- and context-specific, effective guidelines to protect the members of the Profession with less privilege and power, while providing clear actions to take when infractions are suspected or perpetrated. Accountability holds community members responsible for realizing the stated equity-advancing values. The privileging of personal autonomy without systems of accountability has been a major impediment to the realization of ethical, excellent science, a concept that Indigenous Hawaiians refer to as "Imi Pono."[80] Accountability requires that the Profession assess and modify distribution of power, and allow for restorative justice processes that are responsive to the needs of those who have been victimized when considering commensurate consequences.[81] Discriminatory, inequitable, and unethical systems and people must be addressed, including a punitive response for consistent and/or egregious violations of ethical policies. Demonstrated accountability must be structural, data-driven, and site-specific, and include active learning and listening opportunities for all community members. Accountability fosters collective professional well-being, in support of sponsoring agency values, by decisively and quickly addressing infractions when they occur, and supporting structures necessary to reduce the likelihood for such infractions to occur in the first place. This will reduce attrition in the field and increase the vitality of the Profession.

These human-centered, equity-advancing values have been embedded into this report's suggestions, ensuring alignment between the proposed growth of the profession and the guiding

[71] C.Y. Tang and C. Byrge, 2016, Ethnic heterogeneous teams outperform homogeneous teams on well-defined but not ill-defined creative tasks, *Innovation*, 2.

[72] See https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2837428/.

[73] K.W. Phillips, et al., 2014, How diversity works, *Scientific American*, 311(4):42–47.

[74] B. Hofstra, et al., 2020, The diversity–innovation paradox in science, *Proceedings of the National Academy of Sciences*, 117(17):9284–9291.

[75] H.H. Friedman, L.W. Friedman, and C. Leverton, 2016, Increase diversity to boost creativity and enhance problem solving, *Psychosociological Issues in Human Resource Management*, 4(2):7–33.

[76] K.A. Jehn, G.B. Northcraft, and M.A. Neale, 1999, Why differences make a difference: A field study of diversity, conflict and performance in workgroups, *Administrative Science Quarterly*, 44(4):741–763.

[77] T.H. Cox and S. Blake, 1991, Managing cultural diversity: Implications for organizational competitiveness, *Executive*, 5(3):45–56, *JSTOR*, www.jstor.org/stable/4165021, accessed 18 August 2020.

[78] S.A. Hewlett, M. Marshall, and L. Sherbin, 2013, How women drive innovation and growth, *Harvard Business Review–HBR Blog Network*, August 23.

[79] United Nations Framework Convention on Climate Change, 1992 Treaty, 2015, "Common but Differentiated Responsibilities and Respective Capabilities," 30 July, https://climatenexus.org/climate-change-news/common-but-differentiated-responsibilities-and-respective-capabilities-cbdr-rc/, accessed 18 August 2020.

[80] A. Kalili, quoted in B. Lewis, "Plenary Lecture: The Stewardship of Maunakea's Legacy from the Perspective of the Hawaiian and Astronomical Communities," https://astrobites.org/2020/01/07/astrobites-at-aas-235-day-2/, accessed 27 August 2020.

[81] See http://restorativejustice.org/, accessed 18 August 2020.

principles of the federal agencies. The astronomical community will ultimately be responsible for the acceptance and implementation of these values; they cannot be successfully implemented if deployed in a "top down" fashion.[82] *The hope is that the Profession moves forward from the challenges of 2020 with a collective commitment to equity and scientific excellence, a clear plan as laid out in the following suggestions, and benchmarks for progress that can be evaluated in the future.*

N.6 GOALS AND SUGGESTIONS

The panel's overarching goal is to invigorate the Profession through training and workplaces that reflect equity-advancing values and allow the full human diversity of the nation to meaningfully and maximally contribute to the field. The panel's suggestions regarding the state of the Profession are situated within the current landscape and are intended to be actionable within the decade. As is typical for a decadal survey, the primary actors for most of these recommendations are the funding agencies that sponsored Astro2020. However, the panel also includes recommendations for the academic departments, private foundations, observatories, professional societies, government laboratories, and research centers where astronomers work.

In the following sections, the panel presents 7 goals with a total of 18 suggestions considered critical to begin to create a profession that promotes equity-advancing values in order to achieve scientific excellence. Each section follows the same format: first, research findings are presented to motivate each suggestion, and then example methods are provided that can help to carry out the given suggestion, identifying actors, impact, and estimated costs. The findings, suggestions, and methods that are presented in this report are not intended to be prescriptive but may include detail for clarity. Each agency would need to adapt the suggestions to work within their own context to achieve the proposed goals.$

N.6.1 Goal 1: Collecting, Evaluating, and Acting on Demographic Data

Collect and analyze demographic data wherever astronomy research, education, or training is conducted and create internal agencies or society offices to review data and suggest policy change.

To achieve a diverse and inclusive profession requires a robust mechanism to (1) collect data pertinent to the values the Profession espouses, (2) report those data for transparency and accountability, and (3) use the data to compare outcomes to the desired state and adjust as needed. Without data, it is not possible to fully assess the state of the Profession or determine progress toward desired outcomes.

N.6.1.1 Current Practices in the Collection of Demographic Data

The panel requested data on astronomy-related programs from NASA, NSF, DOE, and management organizations for major astronomical facilities. Demographics of staff, contractors, review panels, proposers, awardees of grants and fellowships, and proposal success rates were also requested. Last, the panel sought data on agency programs and funding to promote broader access to opportunities and reduce barriers to achieving success in the field for underrepresented groups.

Minimal data were produced by the federal agencies. While all three agencies collect some demographic data (gender, race, ethnicity) on staff and applicants for funding, several issues are clear. First, the agencies do not collect and track the same quantity or categories of demographic data. NSF collects demographic information, but publishes it only at the highest level of aggregation, and data on

[82] F. Dobbin and A. Kalev, 2018, Why doesn't diversity training work, *Anthropology Now* 10(2):48–55.

people from underrepresented groups are often suppressed to maintain confidentiality.[83,84] After a 2015 critique by the Government Accountability Office,[85] NASA began collecting additional demographic data through its proposal submission website NSPIRES,[86] but the data are not yet publicly available. The DOE Portfolio Analysis and Management System (PAMS)[87] collects applicant demographic data, but it is not designed for analysis, and separate programs in the Office of Science maintain their own databases. Through their diversity-equity-inclusion website, the DOE Office of Science collects and reports demographic data on laboratory employees, although not on facility users.[88]

Second, the agencies have no consistent policy on releasing information. NASA shared the inferred binary gender of awardees based on given names and provided data on the number of unsuccessful proposals in various programs. By contrast, NSF declined to share information of this type, reserving the data it gathers for use in internal reviews and assessments. Third, even if the requested data were collected, it was not readily available, or the panel had to aggregate the information itself. Last, none of the agencies appear to evaluate the efficacy of programs funded to promote diversity and inclusion.

The panel requested demographic data from two management organizations operating major astronomical facilities for NSF: AURA and AUI. Both show commitment to the equity-advancing values of diversity and inclusion and provided demographic data to measure progress. The AIP Statistical Research Center collects longitudinal data and reports on the demographics and career outcomes of students and faculty in both physics and astronomy programs, such as those presented in Section N.3 above. The difficulties encountered by National Academies committees in gathering demographic data from federal agencies are not new.[89] The panel recognizes that the agencies must comply with a number of statutes and regulations governing the collection and release of data[90] such as that requested by the panel.

N.6.1.2 A Cost-Effective Path Forward

An effective path forward is illustrated by the National Institutes of Health (NIH). For decades, it has collected demographic information from researchers in its external grants program, ~80,000 applications/year, larger than the NASA, NSF, and DOE grants programs combined. The Office of Extramural Research manages the process through its electronics grant system, eRA. Applicants submit demographic data on a voluntary basis. As with NASA, NSF, and DOE, these data are not used in the grant decision-making process. However, unlike NASA, NSF, and DOE, the NIH has aggregated and published applicant data on funded programs in its Data Book[91] for decades while maintaining respondent confidentiality. The U.S. Department of Health and Human Services manages an even larger database—RePORTER. It draws information from databases of funded projects and is used by several major federal agencies.[92]

[83] The NSF National Center for Science Engineering Statistics, www.nsf.gov/statistics/about-ncses.cfm#service.
[84] Report on Merit Review, 2019, www.nsf.gov/statistics/about-ncses.cfm#service.
[85] "Women in STEM Research: Better Data and Information Sharing Could Improve Oversight of Federal Grant-Making and Title IX Compliance," https://www.gao.gov/products/GAO-16-14.
[86] See nspires.nasaprs.com/external/.
[87] See www.energy.gov/science/office-science-funding/sc-portfolio-analysis-and-management-system-pams.
[88] See https://www.energy.gov/science/diversity-equity-inclusion.
[89] NAS (National Academy of Sciences), 2000, *Federal Funding of Astronomical Research*, 55, Washington, DC: The National Academies Press, www.nap.edu/read/9954/chapter/8#54.
[90] Notably, the Privacy Act of 1974 (P.L. 93-579) and the Paperwork Reduction Act of 1995 (P.L. 96-511).
[91] See report.nih.gov/nihdatabook/.
[92] See exporter.nih.gov/faq.aspx.

In 2011, working in collaboration with the NIH to obtain nonpublic demographic data on *submitted* grants, Ginther et al.[93] identified a significant gap in African American versus white applicant funding levels at the NIH. This result spawned creation of a Working Group on Diversity. Their 2012 report to the NIH[94] created a new Office for Scientific Workforce Diversity,[95] whose Chief Officer reports directly to the NIH Director, and new funding designed to sustain scientists from underrepresented groups.[96] Early reports on the progress made by the new funding for underrepresented groups show promise.[97]

To better utilize its vast amount of data, NIH plans to contract with an existing federal statistical agency to manage and analyze all of its information, including data on its intramural workforce, composition of review panels, and demographic information on submitted grants, not currently included in the Data Book or RePORTER. The federal statistical agency will have the data management expertise and congressional approval to provide high-level analysis and reports on NIH data to the public (such as the National Center for Health Statistics or the Census Bureau). NIH is poised to make significant change through these initiatives listed in its new Strategic Plan for Workforce Diversity.[98]

Goal 1, Suggestion 1: The panel suggests that federal agencies collect, analyze, and make available demographic, career, and workplace data on members of the Profession.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF, NASA, DOE**
  - *Method:* Following NIH's lead, the agencies can arrange for an existing statistical federal agency to analyze and report their data for them. All existing data could be handed over now, including demographic data, surveys on workplace environment, training grants, and program assessments. Permission can be obtained from the Office of Management and Budget (OMB) to collect any new data. A shared interagency agreement on data collected will ensure that categories and formats are consistent across agencies, follow OMB standards, and allow for benchmarking progress.
  - *Impact:* Existing data and surveys will be publicly available for analysis, allowing for direct feedback on programs and processes. Effective programs can be expanded or emulated elsewhere to increase impact.
  - *Programmatics:* Achieve by 2025. Cost: $700,000/year for each agency's entire portfolio[99] in order to analyze existing data and establish consistent data collection goals across the agencies as a standard procedure.[100]
- **Academic Departments, Non-Federal Institutions, and Professional Societies**
  - *Method:* Following the recommendations of the AAS Task Force on Diversity and Inclusion in *Astronomy* Graduate Education, astronomy departments form a central data

[93] D.K. Ginther, W.T. Schaffer, J. Schnell, B. Masimore, F. Liu, L.L. Haak, and R. Kington, 2011, Race, ethnicity and NIH research awards, *Science*, 333(6045):1015–1019, pubmed.ncbi.nlm.nih.gov/21852498/.

[94] See https://acd.od.nih.gov/documents/reports/DiversityBiomedicalResearchWorkforceReport.pdf.

[95] See https://diversity.nih.gov/.

[96] See https://www.nih.gov/news-events/news-releases/nih-awards-31-million-enhance-diversity-biomedical-research-workforce, accessed 7 November 2020.

[97] See https://diversity.nih.gov/sites/coswd/files/images/docs/ACD_2019_June_13_Valantine_Wilson_FINAL.pdf.

[98] See diversity.nih.gov/sites/coswd/files/images/2018-06/SWD_StrategicPlan_layout_final_links-508c.pdf.

[99] Based on discussions with relevant actors within NIH, such a service will cost NIH, similar in size to NASA and DOE, a nominal fee of a few hundred thousand dollars annually. NSF may cost even less given their relative size.

[100] Based on discussions with relevant actors within the NIH, this would be a one-time cost for the effort.

collection and analysis unit at a relevant professional society to house their demographic and climate data.
- *Impact:* Institutions can collect and store sensitive demographic and climate data without the risk of *violating* respondents' confidentiality and measure progress toward their diversity and inclusion goals.
- *Programmatics:* Achieve by 2025. Cost: $150,000/year.[101]

Goal 1, Suggestion 2: The panel explicitly suggests that each federal agency consider convening a dedicated office to increase oversight, transparency, and accountability. The offices will use data to document progress toward the realization of equity-advancing values and thereby a Profession that evinces inclusion and workforce diversity.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF, NASA, DOE**
  - *Method:* The panel suggests that each agency use this new office to carry out key functions of values-based, equity-advancing management of funds, programs, and assets. This office would regularly update an external advisory board composed of members of the Profession and stakeholders, which could include the Astronomy and Astrophysics Advisory Committee (AAAC).[102] While the specific implementation for each federal agency will depend on its regulatory structure, an example of the methodology an agency might pursue could include: (1) Sponsor annual "town-hall" style events[103] to provide opportunities for members of the Profession to engage with the new agency office. (2) Provide mechanisms for data-driven accountability to ensure that programmatics reflect equity-advancing values that are derived from agency founding documents. For example, by using the demographics gained from partnering with existing federal statistical agencies, the new office can identify and remedy structural inequities in resource allocation and access (including, as it relates to Americans with Disabilities Act [ADA]). (3) Build a set of agency-specific guidelines and procedures that ensure that the Profession engages ethically with all stakeholders.[104]
  - *Impact:* The Profession will gain data-driven insight regarding the realization of values-based equity in the field as evidenced by trends in demographic and other data collected by the agencies.

[101] The panel estimates one FTE society staff person would be needed to manage this effort for all U.S. departments.

[102] See https://www.nsf.gov/mps/ast/aaac/charter.pdf, accessed 24 August 2020.

[103] Ideally, these events would take place at different locations that are representative of the broad array of contexts in which professional astronomy is done, which could be a way to support equitable participation.

[104] The panel has identified a partnership framework, described in the section "Goal 6: Cultivating Local and Global Partnerships," that is gaining traction in scientific and industrial communities. For example, U.S. Endowment for Forestry and Communities, "The Status of Community Based Forestry in the United States," https://www.usendowment.org/the-status-of-community-based-forestry-in-the-united-states/, accessed 24 August 2020; Viswanathan, et al., 2004, "Community-Based Participatory Research: Assessing the Evidence: Summary," *AHRQ Evidence Report Summaries*, https://www.ncbi.nlm.nih.gov/books/NBK11852/, accessed 24 August 2020. A list of internal policies and external resources is provided on *Community Engaged Research* by Ohio State University—Office of Responsible Research Practices, https://orrp.osu.edu/irb/research-participants/community-engaged-research/, accessed 24 August 2020.

— *Programmatics:* To be achieved in 1–5 years. Cost for events: $500,000/year/agency; one-time cost: $250,000/agency.[105] At the time of the National Academies mid-decadal review, sponsoring agencies can have an office in place with a strategic plan and be able to demonstrate progress toward values-based, equity-advancing outcomes and data-driven accountability structures.

N.6.2 Goal 2: Leveraging Power

Leverage funding structures to expand diversity through inclusive workplaces and equitable practices.

Recent National Academies[28] and AIP reports[25] summarize research that shows that science workplace and higher education experiences differ across demographic groups. Individuals with historically marginalized identities report feeling less comfortable than those from dominant groups, and are disproportionately subject to microaggressions, bullying, and harassment to the detriment of their focus and productivity. Conversely, positive mentoring and interpersonal connections built through networking can instill a sense of belonging within the scientific community, ameliorate negative organizational climates, and aid in performance and retention. These findings highlight the significant impact that improving the experience of higher education and the workplace for all scientists can have on excellence in science.

The federal agencies fund basic research nationwide. The organizations that make up the Profession train, promote, and reward astronomers as they advance in their careers. Together the agencies and the Profession can form a powerful partnership to (1) *motivate* the building of equitable and inclusive workplaces and higher education settings; (2) hold each other, as well as members of the Profession *accountable* for engaging in these efforts; and (3) *assess* their progress on this goal.[106]

N.6.2.1 Motivate Individuals to Enhance Effective Mentoring Practices

Individual scientists significantly impact how other community members experience the workplace, most obviously through mentoring.[107] The agencies have uneven requirements for mentoring plans, and the Profession does not adequately train mentors.

Goal 2, Suggestion 1: The panel suggests that federal agencies partner with organizations in the Profession to motivate and support individual PIs to create healthy workplaces, by updating the grants system to require: (1) demonstrated knowledge of evidence-based mentoring practices as well as resources for mentees; (2) reporting and assessment of mentoring built into proposals and reports systems.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF, NASA, DOE**

[105] The panel arrived at this cost by estimating three agency representatives and one administrative staff person for each agency. The one-time allocation would support the development of the ethical guidelines and procedures to the field.

[106] Note that AAS is already partnering with SEA Change, which institutes a reward system to encourage positive culture change. The panel's suggestions focus on using the unique resources and position of federal agencies as an additional avenue for this work.

[107] NASEM (National Academies of Sciences, Engineering, and Medicine), 2019, *The Science of Effective Mentorship in STEMM*, Washington, DC: The National Academies Press.

   - *Method:* (1) Require Mentoring Plans (MPs) in all individual grant proposals that include funding for mentees, whether students, postdoctorates, or staff; (2) utilize NIH's National Research Mentoring Network as an educational resource for evidence-based professional development; and (3) devote 2 percent of grants awarded[108] to work on inclusion or broadening participation.
   - *Impact:* Encourages PIs to enhance mentoring skills through self-assessment, planning, and access to resources; improves mentor/mentee relationships.
   - *Programmatics:* No cost. Could be implemented immediately.
- **NSF, NASA, DOE**
   - *Method:* Make career development of mentees a focus of federally funded programs. (1) Improve graduate programs by (a) giving student stipends to faculty teams within Ph.D. departments charged with reviewing and updating training and mentoring practice, such as NIH's Institutional Predoctoral Training Grants T32[109] and NSF's Research Traineeship Program;[110] (b) engaging institutions to support, assess, and hold accountable research teams to participate in agency-sponsored surveys and assessments of mentoring; and (c) aggregating this data nationwide and longitudinally and reporting to their respective dedicated office to assess the effectiveness of mentoring. (2) Improve development of junior team members through multi-institutional grants, following NASA's Theoretical and Computational Astrophysics Networks model, but requiring explicit mentoring and equity-advancing goals alongside research outcomes.
   - *Impact:* Encourages researchers to work in teams to proactively support junior researchers; builds mentor networks for trainees and channels for information, support, and professional development; encourages establishment of support and accountability for effective mentoring at the institution and within grant structures; collects survey, demographic, and outcome data to assess programs.
   - *Programmatics:* Following NIH, each agency would transfer 10 percent of graduate student support from research grants to individual PIs to team-based programs. Create partnerships with organizations in this work through Institutional Commitment Letters.[111]

N.6.2.2 Incentivize Teams to Support Career Development for Their Members

Current teams leading observational surveys, facilities, and missions do not reflect the diversity in the field in part because diversity considerations are not yet fully incorporated into the funding process. Accessibility in terms of ADA compliance and training in the use of data are not required in data management plans. Agencies do not require reporting of demographics or climate. Diversity of the team is not a consideration in selection, and all-male proposal teams are common.[112]

Equitable access to training is needed for junior scientists to become PIs. For example, there is clear bias in the gender and career stage of leaders and participants in proposals submitted over the past decade to NASA's Explorer-class mission calls. The data suggest that barriers to participation exist in the

---

[108] For a typical individual grant, this would range from a few thousand to several thousand dollars, a scale that encourages deliberate thought about how to spend it—for example, on hiring an undergraduate for a summer.

[109] See https://www.nigms.nih.gov/training/instpredoc.

[110] See https://www.nsf.gov/publications/pub_summ.jsp?WT.z_pims_id=505015&ods_key=nsf19522.

[111] This method comes from the NIH model, where 10 percent of graduate students are supported with training grants.

[112] J. Centrella, M. New, and M. Thompson, 2019, Leadership and participation in NASA's Explorer-class missions, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1909.10314.

development of mission leaders, in the selection of proposals to receive the institutional support required to be competitive, and in a selection process that in practice does not value diverse teams.

> **Goal 2, Suggestion 2:** The panel suggests that federal agencies urge teams (collaborations, projects, facilities, and missions) to adopt evidence-based practices to (1) address demographic disparities in recruitment, retention, and advancement of scientists; (2) provide facilities and data that are accessible to all; (3) implement strategies to improve work environments for all; and (4) assess their own progress.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF, NASA, DOE**
  - *Method:* Expect teams applying for awards to (1) describe plans to demonstrate diversity among members, including technical and leadership; (2) participate in agency-sponsored demographic and climate assessments; (3) have an explicit leadership selection process; (4) have clear mentoring and advising plans for students and postdoctoral fellows; (5) have demonstrated a plan for increasing accessibility for facilities, with open and equitable access to data, software, and training sets; and (6) demonstrate funding and resources devoted to the work above and broadening participation more generally. Approaches would be agency specific. For example, NASA might give extra weight in its selection process to missions with diverse leadership and participation.
  - *Impact:* Sets the expectation for the field; recognizes scientists and their work environments as essential to the development of science itself; aids development of a diverse cohort of future leaders.
  - *Programmatics:* No cost beyond development of assessments.

N.6.2.3 Strengthen Oversight of and Accountability for Funding

The barriers that hinder individuals with historically marginalized identities from choosing and continuing in physics and astronomy, or advancing to positions of power and influence, are elucidated throughout this report. A key barrier is lack of accountability: good mentoring is not trained for or rewarded; there are few consequences for identity-based harassment or bullying; and there is inadequate support for reporters of such problems. The resulting discriminatory loss of talent is unacceptable if the Profession is to maximize innovation and scientific excellence. NIH has made significant efforts to systematically address these problems over the past decade, with the establishment of several groups responsible for the oversight of funded programs, such as the Division of BioMedical Research Workforce and the Diversity Program Consortium. These groups have distinct roles in contributing to the overall agency strategy on workforce development.[113]

> **Goal 2, Suggestion 3:** The panel suggests that DOE, NASA, and NSF build on NIH experiences to strengthen their resources and expertise for education, monitoring, and assessment of proposals and grantees.[114] Accountability policies and processes, tied to proposal rejection or even suspension of funding in extreme cases, would be implemented as part of the funding process.

Method, impact, and programmatics and cost to achieve this suggestion:

---

[113] See https://diversity.nih.gov/sites/coswd/files/images/2018-06/SWD_StrategicPlan_layout_final_links-508c.pdf.

[114] Through the dedicated agency-specific offices suggested in Goal 1.

- **NSF, NASA, DOE, in Partnership with Institutions**
  - *Method:* Use dedicated offices (in Goal 1, Suggestion 2) to support Goal 2, Suggestions 1 and 2. (1) Associated with grant solicitations, provide expectations and resources for the development of several existing components of proposals, including Broader Impacts, Mentoring, Data Management, Facilities Plans, and Institutional Support Letters. For grants to teams, this would further include the development of new plans for self-assessment, leadership, and support of junior members. (2) These components of each proposal would first be reviewed by the agency office and these ratings incorporated during the scientific review process. (3) Proposals would articulate goals and benchmarks outlined in (1), and annual reports would document progress on these to be reviewed by the agency office.
  - *Impact:* Motivates self-education and institutional action toward equitable practices and creating inclusive workplaces. Positive response to a solicitation or a history of effective practices becomes an influential condition for funding. Builds on NIH experience.
  - *Programmatics:* $0.25 million/year/agency office for consultant work to change proposal and annual reporting processes.[115]

N.6.2.4 Increase Funding and Recognition for the People Who Lead the Recruitment, Retention, and Advancement of Individuals from Historically Underrepresented Groups

For those who lead the recruitment, retention, and advancement of individuals from historically underrepresented groups, many of whom are members of historically marginalized groups themselves, this important work can take time and energy that compromises their professional well-being and career.[116,117,118] Grants supporting this work (e.g., NSF S-STEM, REUs) often have rigid funding models that do not acknowledge the loss of scientific productivity of leading PIs or their need for administrative support.

> **Goal 2, Suggestion 4:** The panel suggests that the federal agencies provide material support to researchers who build and lead programs designed to retain, recruit, and advance historically underrepresented people.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF, NASA, DOE**
  - *Method:* Increase budget size and funding flexibility for grants to advance equity-advancing values (e.g., reduce barriers to diversity and equity; create inclusive workplaces) to allow individuals to fund (1) their research program (e.g., pay for graduate students, summer salary, computing resources); and (2) administrative support staff, including program coordinators and evaluators.

---

[115] Cost scaled from $1 million budget for NIH's Division of Biomedical Research Workforce that serves a similar role.

[116] K.B. Porter, J.R. Posselt, K. Reyes, K.E. Slay, and A. Kamimura, 2018, Burdens and benefits of diversity work: Emotion management in STEM doctoral students, *Studies in Graduate and Postdoctoral Education* 9(2).

[117] V. Lerma, L.T. Hamilton, and K. Nielsen, 2020, Racialized equity labor, university appropriation and student resistance, *Social Problems*, 67(2):286–303.

[118] D.R. Hekman, S.K. Johnson, M.-D. Foo, and W. Yang, 2016, Does diversity-valuing behavior result in diminished performance ratings for non-white and female leaders? *Academy of Management Journal*, 60:2.

- *Impact:* Increase the respect for the work it takes to lead and build such programs as well as the ability of scientists to engage in such efforts while maintaining active research programs.
- *Programmatics:* Estimated at $1 million/year/agency to provide extra support for 5–10 grantees (DOE, NASA, NSF).[119]

N.6.3 Goal 3: Reimagining Leadership

Develop, select, and sustain diverse cohorts of leaders who lead by exercising equity-advancing values.

Expanding astrophysical knowledge in the 2020s requires reimagining leadership. The panel envisions a profession that develops and sustains broadly diverse cohorts of leaders who lead by exercising equity-advancing values. Leadership is a social process by which an individual or a group of individuals with a shared vision act to influence, guide, and motivate members of a group to achieve a desired outcome. The Profession currently relies on hierarchical leadership structures that oversee teams to achieve collective research goals.[120] Leaders also oversee the processes that distribute resources, evaluate performance, and recognize scientific excellence. How leaders are cultivated, and how they are encouraged to lead, will determine the advancement of the Profession and the individuals within it.

N.6.3.1 Develop and Select Diverse Leaders Who Practice Equity-Advancing Values

Diverse teams can outperform and out-innovate homogeneous teams.[121] Currently, the absence of an equity-based values framework and the associations of leadership with whiteness, masculinity, and elite education[122] together cause the Profession to preferentially select leaders from over-represented identities and perspectives.[123] These selection processes do not take into account the diversity of skills required to support, advance, and execute the scientific mission. Aspiring leaders are expected to change their leadership styles to conventional norms.[124] Consequently, the Profession's power structure indirectly, but systematically, discriminates and perpetuates the underrepresentation of leaders who lead in diverse ways, including those from historically marginalized groups.[125] Current and future generations of scientists are looking for leaders not only with conventional scientific reputations but also with expertise in the knowledge and skills to combat systemic inequality within the Profession.[126] Therefore, there is an acute need for training leaders with multimodal expertise at all career levels. Such leaders are

---

[119] Cost calculated based on 10 PIs per agency with grants of about $100,000 per year to support their research efforts. This is comparable to current NSF AST spending on REU.

[120] NRC (National Research Council), 2015, *Enhancing the Effectiveness of Team Science,* Washington, DC: The National Academies Press, doi: 10.17226/19007.

[121] V. Hunt, et al., 2018, "Delivering Through Diversity," The McKinsey Report; C. Díaz-García, A. González-Moreno, and F.J. Sáez-Martínez, 2013, Gender diversity within R&D teams: Its impact on radicalness of innovation, *Innovation,* 15(2):149–160, doi: 10.5172/impp.2013.15.2.149; D. Rock and H. Grant, 2016, Why diverse teams are smarter, *Harvard Business Review*; S.S. Levine, et al., 2014, Ethnic diversity deflates price bubbles, *PNAS*, 111(52):18524–18529; doi: 10.1073/pnas.1407301111.

[122] H. Liu, 2018, Redoing and abolishing whiteness in leadership, after Leadership, 101–111; L.A. Rivera, 2016, *Pedigree: How Elite Students Get Elite Jobs*, Princeton, NJ: Princeton University Press.

[123] E. Cech, 2015, Engineers and engineeresses? Self-conceptions and the development of gendered professional wdentities, *Sociological Perspectives,* 58(1):56–77, doi: 10.1177/0731121414556543.

[124] S. Cheryan and H.R. Markus, 2019, Masculine defaults: Identifying and counteracting hidden cultural biases, *Psychology Review*—under review; S.S. Levine, et al., 2014, Ethnic diversity deflates price bubbles, *PNAS*, 111(52):18524–18529, doi: 10.1073/pnas.1407301111.

[125] See https://www.nature.com/articles/d41586-020-01741-7, accessed 26 August 2020.

[126] See https://aas.org/press/aas-endorses-vision-statement-inclusive-astronomy.

defined here as leaders who practice equity-advancing values, including being trained in cultural competency, critical thinking, how to lead discussions inclusively, and how to develop culturally responsible solutions.[127] These skills are hallmarks of multimodal expertise[128] and are essential to leading astronomy in realizing a holistic view of scientific excellence. In addition to new programming, existing leadership training programs in astronomy and physics promote training in the advancement of equity-advancing values (e.g., Project Kaleidoscope,[129] SACNAS Leadership Institute,[130] PI Launchpad,[131] NSBP/NSHP Student Leadership Summit). These excellent models merit financial support, expansion, and replication. There is no need to wait to diversify astronomy's leadership. Effective leaders with multimodal expertise already exist in the Profession and need to be supported to assume greater roles.

Goal 3, Suggestion 1: The panel suggests that members of the Profession purposefully develop, nominate, and select future leaders with multimodal expertise who exercise equity-advancing values. The panel suggests that federal agencies: (1) update selection processes and criteria to require evidence of ability to lead diverse teams; (2) build programs that incentivize the hiring of leaders capable of supporting underrepresented scientists; and (3) develop leadership pathways that include both training in the practice of equity-advancing values and opportunities for early career leadership.

Method, impact, and programmatics and cost to achieve this suggestion:

- **The Profession**
  - *Method:* Update selection criteria for leadership positions throughout the Profession's organizations to include evidence of multimodal expertise through concrete examples where candidates exercise equity-advancing values. Criteria might include demonstrated, quantifiable outcomes—for example, improving institutional culture, building or sustaining effective community partnerships,[132] demonstrating academic leadership on these topics (publications, lectures, and discussions), and improving recruitment, retention, and advancement of mentees, particularly individuals from historically underrepresented communities.
  - *Impact:* Reduce current inequities in access to resources, awards, advancement, and leadership appointments through the selection of leaders who practice equity-advancing values. Select leaders who have the skills needed to support a diverse workforce.
  - *Programmatics:* No-cost. Can be implemented immediately.[133]
- **DOE, NASA, NSF, Academic Institutions, Government Laboratories/Observatories**
  - *Method:* Diversify institutions' permanent professional workforces with respect to race/ethnicity/gender and other social identities.[28] The panel suggests that institutions and agencies build hiring programs to incentivize the creation of new positions for individuals

---

[127] S. Lee, 2020, *Yale Astronomers Questioned Systemic Racism Because They Hired One Black Employee 35 Years Ago, Emails Show*, Buzzfeed, https://www.buzzfeednews.com/article/stephaniemlee/yale-astronomy-systemic-racism-emails, accessed 24 August 2020.

[128] J. Alvehus, 2019, Emergent, distributed, and orchestrated: Understanding leadership through frame analysis, *Leadership*, 15(5):535–554, doi:10.1177/1742715018773832. See also the Section N.5 "Values Statement" in this document. W. Kuepers, 2012, Donna Ladkin, Rethinking leadership: A new look at old leadership questions, *Leadership*, 8:463–467, doi: 10.1177/1742715012444678.

[129] Project Kaleidoscope (Leadership in STEM Training), https://www.aacu.org/summerinstitutes/sli/2018, accessed 24 August 2020.

[130] SACNAS Leadership Institute, https://www.sacnas.org/what-we-do/leadership-programs, accessed 24 August 2020.

[131] PI Launchpad, https://science.nasa.gov/researchers/pi-launchpad, accessed 24 August 2020.

[132] See the Section "Cultivating Local and Global Partnerships" in this document.

[133] Requires only additional criteria in selection procedures for leadership positions and awards.

with strong track records in promoting equity-advancing values. Agencies could follow solicitation NSF 19-558, Faculty *Development in the Space Sciences*. The panel further suggests that agencies and institutions classify the recruitment of such employees as critical, with a severe shortage of candidates, and support this priority by utilizing tools at their disposal, such as the Direct Hiring Authority.[134]
  - *Impact:* Science leaders with demonstrated equity-advancing skills who support scientists from underrepresented backgrounds.
  - *Programmatics:* Estimated at $4.5 million per agency (DOE, NASA, NSF).[135]
- **DOE, NASA, NSF**
  - *Method:* Build leadership training programs specific to the agency's leadership structures and include workshops to teach how to implement equity-advancing values as leaders. For example, missions and collaborations might include leadership development in their budgets, and each agency establishes and funds equivalent to the PI Launchpad program. The panel suggests that outcomes from training programs be assessed with longitudinal tracking of participants and reporting of aggregated data.
  - *Impact:* Agencies participate and guide the development of leadership programs that provide equitable access to organization-specific information.
  - *Programmatics:* Estimated at $120,000 per meeting, per agency.[136]

N.6.3.2 Promote the Exercise of Leadership by Diverse Leaders

STEM organizations have become more diverse primarily through the disproportionate labor of scientists who represent the communities that STEM fields are seeking to better serve.[137] Individuals with historically underrepresented identities spend significant time on this "invisible" work, with consequences to their research productivity.[138] True commitment to exercising equity-advancing values must not obscure the racial equity labor that goes into building racial inclusion.[139] Furthermore, leadership by white women and members of marginalized groups is often unduly scrutinized and criticized.[140] The Profession

[134] Direct Hiring Authority, https://www.opm.gov/policy-data-oversight/hiring-information/direct-hire-authority/, accessed 24 August 2020.

[135] Annual; estimates are based on NSF 19-558, *Faculty Development in the Space Sciences*. Funding supports 3–4 awards per agency, resulting in 9–12 new hires annually. See https://www.nsf.gov/pubs/2019/nsf19558/nsf19558.htm, accessed 24 August 2020.

[136] Annual; estimates are based on the budget for the NASA PI Launchpad program (E. Hamden, private communication): 40 people attending, plus ~25 mentors/speakers/panelists = $100,000 operations, $20,000 travel budget for NASA speakers = $120,000. Budget for PI Launchpad was largely supported by the Heising-Simons Foundation.

[137] J. Posselt, 2020, *Equity in Science: Representation, Culture, and the Dynamics of Change in Graduate Education*, Palo Alto, CA: Stanford University Press; K.B. Porter, J.R. Posselt, K. Reyes, K.E. Slay, and A. Kamimura, 2018, Burdens and benefits of diversity work: Emotion management in STEM doctoral students, *Studies in Graduate and Postdoctoral Education*.

[138] Brown-Nagin, 2016, The mentoring gap, commentary, *Harvard Law Review*, 303:129; C.T. Pittman, 2010, Race and gender oppression in the classroom: The experiences of women faculty of color with white male students, *Teaching Sociology*, 38(3):183–196, doi: 10.1177/0092055X10370120; D.R. Hekman, S.K. Johnson, M.-D. Foo, and W. Yang, 2016, Does diversity-valuing behavior result in diminished performance ratings for non-white and female leaders? *Academy of Management Journal*, 60:2; R.F. Martell, 1991, Sex bias at work: The effects of attentional and memory demands on performance ratings for men and women, *Journal of Applied Social Psychology*, 21:1939–1960.

[139] V. Lerma, L.T. Hamilton, and K. Nielsen, 2020, Racialized equity labor, university appropriation and student resistance, *Social Problems*, 67:2, doi: 10.1093/socpro/spz011.

[140] M.E. Heilman, A.S. Wallen, D. Fuchs, and M.M. Tamkins, 2004, Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89:3; D. Hekman and M.-D. Foo, 2017,

needs to accept and empower leaders with multi-modal expertise by recognizing the value of diverse ways of leading,[141] and be willing to be led by people with different ideas, identities, and approaches.

Recognition is one of the core tenets of belonging, critical to the creation of a STEM identity,[142] and a key determinant of retention.[143] Agencies and the Profession can use powerful levers (awards, grants, prizes, promotion, raises, tenure) to recognize the currently invisible labor of individuals to diversify the Profession. Such levers can help sustain leaders with multimodal expertise who are critical to actualizing equity-advancing values and the strategic plans of agencies/institutions. This establishes the work of promoting equity-advancing values as a core mission of the Profession and a responsibility of its leaders.[144]

> **Goal 3, Suggestion 2:** The panel suggests that the Profession sustain and empower leaders with multi-modal expertise, including leaders from historically underrepresented groups, by recognizing their leadership in encouraging equity-advancing values in promotion evaluation and service assignments. This responsibility lies not only with those who select leaders, but also with their peers and those being led.

Method, impact, and programmatics and cost to achieve this suggestion:

- **The Profession**
  - *Method:* Recognize and reward leadership that demonstrates equity-advancing values in individual evaluations at all career stages—for example, fellowship applications, awards and review committees, evaluation for tenure and promotion. Account for this leadership when considering service loads within institutions so that scientists from historically underrepresented backgrounds (including women) are not overburdened. Provide meaningful, context-specific rewards for scientists who promote equity-advancing values, which can include service/teaching relief and/or an extra semester of sabbatical.
  - *Impact:* Rewarding such leadership influences promotion metrics used at all institutional levels, empowers individuals, particularly those from underrepresented communities, to continue promoting equity-advancing values in the Profession, and encourages others to join in and respect the work of these individuals.
  - *Programmatics:* Minimal up-front cost that is ultimately recoverable.[145]
- **DOE, NSF, NASA**
  - *Method:* Establish Early-Career Leadership Awards and Fellowships to recognize and fund early-career faculty, scientists, postdoctorates, graduate, and especially undergraduate students that work to support the recruitment and retention of historically underrepresented scholars. Create leadership training programs for awardees and existing agency postdoctoral fellows. Self-nominations for awards ought to be encouraged.

Does valuing diversity result in worse performance ratings for minority and female leaders? *Academy of Management Annual Meeting Proceedings 2014*; S.K. Johnson, and D.R. Hekman, 2016, Women and minorities penalized for promoting diversity, *Harvard Business Review*.

[141] L. Madhlangobe and S.P. Gordon, 2012, Culturally responsive leadership in a diverse school: A case study of a high school leader, *NASSP Bulletin*, 96(3):177–202, doi: 10.1177/0192636512450909.

[142] H.B. Carlone and A. Johnson, 2007, Understanding the science experiences of successful women of color: Science identity as an analytic lens, *J. Res. Sci. Teach.*, 44:1187–1218, doi:10.1002/tea.20237.

[143] J.E. Stets, P.S. Brenner, P.J. Burke, and R.T. Serpe, 2017, The science identity and entering a science occupation, *Social Science Research*, 64:1–14. doi: 10.1016/j.ssresearch.2016.10.016.

[144] W. Brown-Glaude, ed., 2009, *Doing Diversity in Higher Education: Faculty Leaders Share Challenges and Strategies*, New Brunswick, NJ: Rutgers University Press.

[145] Requires additional criteria in promotion/selection criteria (no cost). Teaching relief and/or extensions in sabbatical are short-term costs for the institution employing the individuals that can be balanced in the long run by retention and improved performance of employees who improve the climate of the institution.

- *Impact:* Encourages institutions to recognize these individuals and their work to further equity-advancing values.
- *Programmatics:* Financial support for the individual's research through multi-year awards to scientists (similar to NSF/DOE Early Career Awards), fellowships for graduate students and postdoctorates, and scholarship awards for tuition for undergraduates. Award recipients could receive mentoring from dedicated agency-led leadership training programs (Goal 3, Suggestion 1). Estimated at $3 million/year NASA, $3 million/year NSF, $1.5 million/year DOE.[146]

N.6.4 Goal 4: Addressing Harassment and Discrimination

Establish clear policies, collect and report relevant metrics, and enforce accountability measures to remove structures and individuals that perpetrate identity-based discrimination including harassment.

Identity-based discrimination is a core mechanism for preserving inequity within the Profession.[147] It includes both differential treatment (including harassment) on the basis of identity, as well as ostensibly neutral practices that produce differential impacts owing to identity. Identity-based discrimination minimizes equitable access to the resources, infrastructure, and relationships necessary to participate fully in the field, and discourages multimodal expertise by subordinating those historically perceived to be from social out-groups.[148] It erodes the sense of belonging and respect needed for confident engagement, thereby diminishing or altogether eliminating people and their valuable perspectives.[149] Given the pervasiveness of identity-based discrimination (including harassment) in the Profession,[150] the panel emphasizes the need to balance accountability, recourse/reporting and environmental interventions to address and ultimately eradicate unchecked acts of discrimination as well as the standard operating procedures that have disparate or differential impact on individuals in the field.

Pervasive identity-based discrimination in the Profession (be it structural or between individuals, overt or implicit) impacts (1) professional well-being by producing stress and other negative health outcomes; (2) equitable participation and advancement by not accounting for these differences in experience and mental/emotional load when evaluating performance and outcomes; and (3) economic prosperity and innovation by limiting the degree to which minoritized populations can obtain and maintain jobs in the Profession and further a deeper understanding of the universe.

Since 2018, the National Academies have released four consensus reports that have taken a systemic approach in addressing key issues in higher education and academic research: *Graduate STEM*

---

[146] Annual; estimates based on the following: (1) Scientists: Comparable to NSF CAREER, DOE Early Career Programs (5-year term, 500,000 grants, 6 per year; NASA, DOE, NSF). (2) Graduate/Postdoctoral: Comparable to AAPF/GRFP (~$100,000 per fellow, selecting ~15 new fellows per year, for 3-year terms; NASA, NSF). (3) Funding for scholarships for undergraduates ($15,000 per student, 20 students per year, NASA, NSF). Estimate based on data from 2015–2016, "where 78 percent of full-time students at public 4-year colleges and universities had need remaining after grant aid, averaging $14,400." Trends in Student Aid 2019, College Board, https://research.collegeboard.org/pdf/trends-student-aid-2019-full-report.pdf, accessed 24 August 2020.

[147] NASEM (National Academies of Sciences, Engineering, and Medicine), 2020, *Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors*, Washington, DC: The National Academies Press, https://doi.org/10.17226/25585.

[148] K. Crenshaw, 1989, Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics, *University of Chicago Legal Forum*, 1989(1):8.

[149] X. Padamsee, 2017, Unrealized impact: The case for diversity, equity, and inclusion, *Promise54*, July: 52–53.

[150] K.B.H. Clancy, K.M.N. Lee, E.M. Rodgers, C. Richey, 2017, Double jeopardy in astronomy and planetary science: Women of color face greater risks of gendered and racial harassment, *JGR Planets*, 122:1610, https://doi.org/10.1002/2017JE005256.

Education for the 21st Century; *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*; *The Science of Effective Mentorship in STEMM*; and *Minority Serving Institutions: America's Underutilized Resource for Strengthening the STEM Workforce*. Each of the committees created reports that situated the issue of sexual harassment and discrimination within the broader culture of higher education, as the committees perceived that incentive and reward systems are critical drivers of behavior in academia. In particular, there is broad consensus that the legal system alone is not an adequate mechanism for reducing or preventing sexual harassment. These reports further highlight the role that federal agencies, which control research funding, can play in enacting long-lasting change.

The National Academies report *Sexual Harassment of Women*[151,152] highlights the need to address the effects of harassment and discrimination on the integrity of research. This report concludes that "parts of the federal government and several professional societies…focus more broadly on policies about research integrity and on codes of ethics, rather than on the narrow definition of research misconduct."[153] The panel is in agreement that scientific integrity has to include how researchers treat people. "Research culture and policies are quick to denounce plagiarism, data fabrication, and mismanagement of funds, yet we have too long ignored the mistreatment of people."[154] The House of Representatives Committee on Space, Science, and Technology in 2019 held a hearing[155] to investigate efforts to combat sexual harassment in STEM fields. In her opening statement, Chair Eddie Bernice Johnson said, "The public investment in research needs to draw on all of our nation's talent to return the best possible science for the benefit of society. To reach this goal, we must do more to ensure that all researchers have access to a safe work environment." "Harassment, bullying, and discrimination damage science at the individual, community, institutional, and societal levels and cause health problems, fear, mistrust, depression, and trauma."[156] It thus follows that additional consideration needs to be given to safe social spaces, termed "counterspaces," which provide support and reinforce the sense of belonging in STEM.[157] Counterspaces[158] can enable peer-to-peer relationships that provide academic, social, and/or emotional support, mentoring relationships that help victims navigate how to succeed in the field, and access to campus groups to advance professional skills and develop leadership opportunities.[159] Support programs can take the form of coaching, counseling, and childcare while negotiating the after-effects.

Cultural shifts around identity-based harassment require second-order theories of change (i.e., addressing underlying priorities and norms, not just reforming policy and practice) and an intersectional

---

[151] NASEM (National Academies of Sciences, Engineering, and Medicine), 2018, *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*, Washington, DC: The National Academies Press, doi: https://doi.org/10.17226/24994 (Chapter V.5 and Chapter V.6 and R:13).

[152] Page 170 of that report defines sexual harassment; the panel uses that definition.

[153] NASEM (National Academies of Sciences, Engineering, and Medicine), 2018, Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine, Washington, DC: The National Academies Press, doi: https://doi.org/10.17226/24994 (Conclusion 6).

[154] E. Marín-Spiotta, 2018, Harassment should count as scientific misconduct, *Nature*, 557:141.

[155] See https://science.house.gov/hearings/combating-sexual-harassment-in-science.

[156] E. Marín-Spiotta, 2018, Harassment should count as scientific misconduct, *Nature*, 557:141.

[157] D. Solorzano, M. Ceja, and T. Yosso, 2000, Critical race theory, racial microaggressions, and campus racial climate: The experiences of African American college students, *Journal of Negro Education*, 69(1/2):60.

[158] "Counterspaces in science, technology, engineering, and mathematics (STEM) education are often considered safe spaces that, by definition, lie in the margins, outside of mainstream educational spaces, and are occupied by members of non-traditional groups." From M. Ong, J.M. Smith, and L.T. Ko, 2018, Counterspaces for women of color in STEM higher education: Marginal and central spaces for persistence and success, *J. Res. Sci. Teach*., 55:206–245.

[159] NASEM (National Academies of Sciences, Engineering, and Medicine), 2020, *Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors*, Washington, DC: The National Academies Press, https://doi.org/10.17226/25585.

lens (i.e., attending to experiences of people with multiple marginalized identities).[160] Harassment continues to be a major concern in our field. In the most recent poll by the Pew Research Center, "among women who worked in male-dominated workplaces, 48 percent said that harassment was a problem. Just under one-quarter of women said that they had been harassed."[161] In spite of the face that research on gender inequalities in STEM has generated ample strategies in order to achieve gender equity,[162] urgent gaps persist in knowledge about racial discrimination, including how it manifests in educational and professional environments and how it intersects with other forms of discrimination and oppression.[163] These themes necessitate a great deal of reflection and require an intersectional approach. These include, for example: the experiences of women of color, women with disabilities, LGBTQIA+ women, as well as those involving women in all intersectional identities.

Goal 4, Suggestion 1: Recognize identity-based discrimination and harassment as equally deleterious as research misconduct in terms of its effects on the integrity of research.[164]

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF, NASA, and DOE**
  - *Method:* The panel suggests that agencies adopt scientific integrity policies that specifically address identity-based harassment with the same severity as any other research or scientific misconduct, such as fabrication, falsification, or plagiarism. The panel endorses the suggestions of Zellner and collaborators and supports "the provisions of H.R. 36, the Combating Sexual Harassment in Science Act of 2019… The proposed law requires the development and implementation of harassment reporting terms and conditions, like the one used by NSF, at all major science funding agencies."[165]
  - *Impact:* The panel identified grantmaking authorities as the optimal actors. Adding new terms and conditions directed specifically at harassment and discrimination to the agencies' proposal policies would place it alongside numerous other requirements that institutions already agree to every year when they accept funding.
  - *Programmatics:* No-cost. Could be implemented in 1–2 years.
- **NSF, NASA, and DOE**
  - *Method:* Hold individuals and institutions responsible for harassment and discrimination. Establishing sexual harassment as a serious issue would require that federal funding agencies be notified by funded institutions when principal investigators, co-principal investigators, and grant personnel have violated sexual harassment policies.[166]

[160] S. Elrod, and A. Kezar, 2016, *Increasing Student Success in STEM: A Guide to Systemic Institutional Change*, Washington, DC: Association for American Colleges and Universities.

[161] Race- and gender-based bias persists in US science, 2018, *Nature*, 554:561. doi: https://doi.org/10.1038/d41586-018-02175-y

[162] T. Feder, 2017, 2017, Widespread harassment reported in astronomer survey, *Physics Today in Politics and Policy* 21(July).

[163] C. Prescod-Weinstein, 2020, Making Black women scientists under white empiricism: The racialization of epistemology in physics, *Journal of Women in Culture and Society*, 45(2).

[164] "Related Findings and Suggestions," *Sexual Harassment of Women*, Chapter V and Recommendations 3, 4, 13, and 14,

[165] N. Zellner, J. McBride, N. Morrison, A. Olmstead, M. Patterson, G. Rudnick, A. Venkatesan, et al., 2019, "Findings and Recommendations from the American Astronomical Society (AAS) Committee on the Status of Women In Astronomy: Towards Eliminating Harassment in Astronomy," white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1908.00589.

[166] "Findings and Conclusions," *Sexual Harassment of Women*, Chapter V, Number 5.

- — *Impact:* Increases incentives for institutions. Creates accountability partnerships between agencies and institutions; sets expectations of accountability and consequences throughout the field.
- — *Programmatics:* No cost. Can be implemented in 1–2 years.

- **Academic Institutions**
  - — *Method:* The panel suggests that academic institutions consider identity-based discrimination as equally important as research misconduct and increase collaboration among offices that oversee the integrity of research in order to provide more resources to handle complaints and implement sanctions.[167]
  - — *Impact:* By enforcing consequences for identity-based discrimination as a violation of research integrity, institutions can be better equipped to remove individuals and systemic structures that perpetrate identity-based discrimination.
  - — *Programmatics:* No-cost. Can begin in the first year.
- **Professional Societies**
  - — *Method:* The panel suggests that professional societies seek to eliminate harassment and discrimination in their activities, particularly conferences and scientific publication, and throughout the profession by providing resources and setting high community-based standards of conduct.[168]
  - — *Impact:* Would lower the tolerance for harassment and discrimination within the Profession, and promote grass-roots changes in behavior.
  - — *Programmatics:* No-cost. Can be implemented immediately.

Goal 4, Suggestion 2: Support individuals marginalized by harassment and discrimination.[169]

Method, impact, and programmatics and cost to achieve this suggestion:

- **Professional Societies and Private Foundations**
  - — *Method:* The panel suggests that professional societies and private foundations convene working groups that can effectively assess how funding can be provided for mental health and well-being, legal counseling, and other support structures for survivors.
  - — *Impact:* Individuals impacted by discriminatory practices or harassment require a range of support options that can be facilitated by flexible funding that allows them to make arrangements that best suit their needs.
  - — *Programmatics:* Convened working groups could include participation by representatives of funding agencies. Options might include support for dependents or caregivers or for new and existing counterspaces designed to mitigate the negative impacts of identity-based discrimination.
- **Academic Institutions**
  - — *Method:* The panel suggests that academic institutions support new and existing counterspaces designed to mitigate the negative impacts of identity-based discrimination.
  - — *Impact:* "Creating counterspaces, alongside inclusive policies that guard against racism and sexism (and other forms of discrimination), [enhances the] learning environment and the opportunity for all to succeed."[170]

---

[167] *The Sexual Harassment of Women,* Recommendation 4.

[168] S. Sardelis, S. Oester, and M. Liboiron, 2017, Ten strategies to reduce gender inequality at scientific conferences, *Frontiers in Marine Science*, 4:231.

[169] "Related Findings and Suggestions," *The Sexual Harassment of Women*, Recommendations 4, 13, and 14.

[170] M. Ong, J.M. Smith, and L.T. Ko, 2018, Counterspaces for women of color in STEM higher education: Marginal and central spaces for persistence and success. *J. Res. Scie. Teach*., 55:206-245.

- — *Programmatics:* Could be done as an institutional program in campuses around the country.
- **NSF, NASA, DOE, and Academic Institutions**
  - — *Method:* The panel suggests that agencies and institutions design and fund training that focuses on cultural humility and bystander-intervention. While cultural competency focuses on providing practitioners the ability to understand, communicate with, and effectively interact with people across cultures,[171] cultural humility is a way of being with ourselves, others, and the institutions we inhabit.[172] It asks not only that we assess our environments and engage them in an unbiased and nonviolent manner but also that we reflect deeply on who we are and how we show up for others.
  - — *Impact:* Changes discriminatory evaluation and decision-making processes within the Profession through training to reduce inequities in participation and leadership within the field.
  - — *Programmatics:* Minimal cost. Could be implemented in 1–2 years.
- **NSF, NASA, and DOE, and Institutions**
  - — *Method:* Because lack of access is a form of discrimination, the panel suggests that institutions consider developing accessibility plans to identify the current state of facilities and plans for increasing access.
  - — *Impact:* Accessible spaces encourage equal participation.
  - — *Programmatics:* Accessibility plans can be implemented in 1–2 years.

N.6.5 Goal 5: Removing Barriers

Modernize practices that have disparate impact on access to education, training, and advancement.

Scientific excellence depends on ensuring that each generation of scientists can thrive within the environments in which they learn and work, and requires equitable access to education, advancement opportunities, funding, and facilities. Astronomy is a dynamic field, both culturally and technologically, and training (including teaching practices, curriculum, and technical/professional development) that reflects the current state of evidence-based, inclusive practice is needed. Physics and mathematics instruction is the gateway to the Profession and must be modernized nationwide. Inequities in career advancement and access to the tools of the Profession must be addressed so that the entire workforce is engaged. See also the driving motivation for SEA Change,[173] an effort of the American Association for the Advancement of Science to effect sustainable change with regard to diversity, equity, and inclusion in STEMM[174] at U.S. institutions of higher education.

N.6.5.1 Work with Physics Departments to Incentivize the Widespread Adoption of Research-Based Instructional Strategies and Inclusive Pedagogy in First-Year Physics

The first-year sequence in physics is among the most influential in a student's chances to continue not only in astronomy but also in all STEM fields. This sequence, along with calculus, have drop, fail, or withdrawal (DFW) rates of 30 percent or more, and first-generation (First Gen), Pell-eligible

[171] T.L. Cross, B.J. Bazron, K.W. Dennis, and M.R. Isaacs, 1989, Towards a Culturally Competent System of Care, *NCJRS*: 12439.

[172] R. Danso, 2018, Cultural competence and cultural humility: A critical reflection on key cultural diversity concepts, *Journal of Social Work*, 18(4):410–430, doi:10.1177/1468017316654341.

[173] See https://seachange.aaas.org/.

[174] Science, technology, engineering, mathematics, and medicine (STEMM).

and minoritized students, particularly those with intersectional identities, can have nearly double the DFW rates of majority non-Pell, non-First Gen students.[175] Retention and degree completion is strongly tied to D and F grades in the first term.[176] The increased DFW rate at the course level with URM students leads to the loss of URM students presented in the earlier sections. Physics Education Research and Astronomy Education Research (PER, AER) shows that there are specific instructional practices that consistently achieve better student course outcomes and retention than traditional lectures.[177] Collectively known as "interactive engagement," these methods include student-centered instruction and discovery-based learning practices such as peer instruction.[178] Sociology and psychology research further demonstrates the importance of student belonging and the impact of stereotype threat, and provides proven classroom methods that improve student performance.[179] The low rates at which these methods are applied in STEM courses reduces the production of science and technology graduates, and contributes to the loss of diversity among those that do graduate. Moreover, the documented reduction of gender and racial/ethnic performance gaps[180,181] in courses taught with Research-Based Instructional Strategies (RBIS) makes the continued use of lecture-based teaching in first-year physics and calculus courses tantamount to discriminatory practices.

Sadly, evidence abounds that despite efforts to train faculty to move from teacher-centered lecture to learner-centered course design, the majority of faculty trained (75 percent) in workshops continue to use lecture-based pedagogy.[182] Recently, new initiatives that promote the use of RBIS are grounded in robust theories of change, such as supporting networks or learning communities of faculty, called Communities of Practice and Research-Practice Partnerships. Learning communities allow cultural and work-related shifts to happen on the part of both researchers and practitioners engaging in this work to implement and spread reform.[183] A non-exhaustive list includes the Accelerating Systemic Change Network,[184] AAC&U TIDES,[185] the AAU Undergraduate STEM education Initiative,[186] and Kaleidoscope.[187] Private foundations have supported the advancement of such communities, such as the Research Corporation for Science Advancement's contributions to the American Physical Society's NSF Funded Workshop for New Physics and Astronomy Faculty. New funding from federal agencies is required to implement these new, innovative means for increasing the adoption of RBIS. This will require enriched engagement with education researchers in designing professional and department-level training and mentoring in RBIS.

Goal 5, Suggestion 1: The panel suggests that the Profession adopt and promote inclusive pedagogy and RBIS in the classroom through engagement with experts from the PER and AER

[175] See https://www.aplu.org/library/powered-by-publics-learning-memo-the-big-ten-academic-alliance-cluster/file.

[176] [176] See https://www.aplu.org/library/powered-by-publics-learning-memo-the-big-ten-academic-alliance-cluster/file.

[177] Freeman et al., 2014, *PNAS*, 111(23):8410–8415.

[178] Turpen et al., 2016, *Phys. Rev. Phys. Educ. Res.*, 12:010116.

[179] C. Verschelden, 2017, *Bandwidth Recovery: Helping Students Reclaim Cognitive Resources Lost to Poverty, Racism, and Social Marginalization*, Stylus Publishing.

[180] Eddy and Hogan, 2017, *CBE-Life Sciences Education*, 13:3.

[181] Theobald, et al., 2020, *PNAS*, 117:6476.

[182] Ebert-May, et al., 2011, *What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs*, *BioScience*, 61:550–558, doi: 10.1525/bio.2011.61.7.9.

[183] Kezar et al., 2015, Implicit theories of change as a barrier to change on college campuses: An examination of STEM reform, *Review of Higher Education*, 38(4):479–506, doi: 10.1353/rhe.2015.0026.

[184] Accelerating Systemic Change Network, https://ascnhighered.org/index.html.

[185] AAC&U TIDES, https://www.aacu.org/2021-TIDES.

[186] The AAU Undergraduate STEM Education Initiative, https://www.aau.edu/education-community-impact/undergraduate-education/undergraduate-stem-education-initiative.

[187] Kaleidoscope, https://www.aacu.org/pkal.

community to design professional and department-level training in modern teaching practices at all career stages. To achieve transformational change at a national scale, the panel suggests that federal agencies increase funding in PER and AER.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF-MPS, NASA STEM-Engagement**
  — *Method:* Expand funding for research-practice partnerships based on Physics and Astronomy Education Research in order to promote the adoption of evidence-based inclusive pedagogy. Funding supports grants for conferences, training for current and future instructors (master's, doctoral students, and postdoctorates). Private foundations can also support program development.
  — *Impact:* Expanded use of RBIS and inclusive pedagogy in gateway courses will increase retention of all students pursuing astrophysics, particularly underrepresented students.
  — *Programmatics:* $3 million/year NSF-MPS; $2 million/year NASA STEM-Engage.[188]

N.6.5.2 Invest in Programs and Practices to Increase Inclusion and Persistence of Scientists from Groups Historically Underrepresented

Federal funding has created multiple programs to recruit, retain, and advance historically underrepresented people within the Profession, including Bridge Programs (Fisk-Vanderbilt, Columbia, Cal-Bridge, IGEN Bridge); terminal master's programs (e.g., Wesleyan); and summer research programs (REUs, CAMPARE[189]). NSF is funding APS and AAPT's new program, Effective Practices for Physics Programs (EP3) for responding to challenges and engaging in systematic improvements.[190] DOE is funding a Visiting Faculty Program[191] (VFP, formerly known as FaST) to increase faculty and students at institutions historically underrepresented in research areas important to DOE. Last, the National Society of Black Physicists (NSBP), funded by NASA, NIST, NSF, and several national and private research institutions and organizations, has a growing list of student chapters. Such programs and organizations enhance access to doctoral education, as well as a sense of belonging and identity for physics students from underrepresented groups, which increase their persistence and success. However, agencies no longer offer funding for long-term sustainability nor institutional or agency accountability for the continuation of past successful programs. For example, PAARE[192] and MUCERPI[193] are no longer receiving proposals. Investments for programs that have shown progress in increasing the persistence of historically underrepresented groups are most successful if they are not time-limited but are supported for as long as they are effective.

In addition to support for such programs, there is a clear need to remove racial, gender, and other barriers to doctoral education in astronomy and physics, including those created through predominant admissions practices to doctoral education. For example, misuse of the general Graduate Record Exam (GRE) and physics subject GRE (PGRE) in admissions decisions leads to disproportionate exclusion of

[188] Fund 30 groups per year to use complex theories of change to train instructors in RBIS and inclusive pedagogy.
[189] See https://www.cpp.edu/calbridge/summer-research.shtml.
[190] Effective Practices for Physics Programs, https://ep3guide.org/.
[191] DOE VFP, https://science.osti.gov/wdts/vfp.
[192] K.G. Stassun, 2011, The Fisk-Vanderbilt Master's-to-Ph.D. Bridge Program, *American Journal of Physics*, 79:374.
[193] P.J. Sakimoto, and J.D. Rosenthal, 2005, *Physics Today*, September:49–53. [[RSO: PLEASE PROVIDE FULL CITATIONS FOR 181, 184, 185, 186]]

scholars from underrepresented groups, especially women of color.[194,195] These tests have large score gaps by race and gender identities of test takers, yet evidence shows that high scores do not help students to stand out in admission, only penalize otherwise competitive applicants.[196] Further, PGRE scores are not correlated with Ph.D. degree completion[197] nor do they foretell postdoctoral success.[198] Increasingly, astronomy Ph.D.-granting programs are removing GRE and PGRE requirements with no reported negative impact on the academic success of the admitted students.[199]

Goal 5, Suggestion 2: The panel suggests the Profession remove barriers that impede student advancement and renew funding of previous programs with a strong record of retention and advancement of individuals from underrepresented groups.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF-PHY, -AST**
  - *Method:* Provide new (or renewed) funds for programs that recruit, retain, and advance historically underrepresented people to support entry into the Profession. Review impact and internal processes from past funded programs (e.g., PAARE, Fisk-Vanderbilt) to determine if their record in advancing individuals from underrepresented groups merits refunding and/or refinement.
  - *Impact:* Increase program longevity and sustain PI commitment.
  - *Programmatics:* $3 million per year to fund nine sites.[200]
- **Academic Departments**
  - *Method:* Provide funds to reduce or eliminate application fees for low-income and historically marginalized applicants. Eliminate requirements for the GRE and PGRE in admissions to astronomy and physics graduate programs.[201] Replace the traditional admissions process with one that embodies the ideals of equity-advancing holistic review.[202]
  - *Impact:* Increase diversity in graduate programs.
  - *Programmatics:* Marginal department cost of effort to devise and implement holistic admissions process and cover application fees for targeted individuals.

N.6.5.3 Provide Broader Opportunity and Continual Training in State-of-the-Art Techniques

To ensure innovation at an emergent level, technical training programs in computational methods and instrumentation are needed for astronomers throughout their careers. Computational and data

194 C. Miller and K. Stassun, 2014, A test that fails, *Nature*, 510:303–304.

195 J. Posselt, 2016, *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*, Cambridge, MA: Harvard University Press.

196 N.T. Young and M.D. Caballero, 2020, The physics GRE does not help applicants “stand out,” arXiv:2008.10712.

197 Miller, et al., 2019, *Science Advances*, 5(1).

198 E.M. Levesque, E.M. et al., 2015, *Physics GRE Scores of Prize Postdoctoral Fellows in Astronomy*, arXiv:1512.03709.

199 Due in part to COVID-19, GRE and pGRE were eliminated from most admission requirements this cycle.

200 Estimates based on previous PAARE funding ($1 million/site for 3 years).

201 Also recommended by numerous previous reports, most recently the Nashville Recommendations: https://tiki.aas.org/tiki-index.php?page=Inclusive_Astronomy_The_Nashville_Recommendations.

202 J.D. Kent and M.T. McCarthy, 2016 *Holistic Review in Graduate Admissions: A Report from the Council of Graduate Schools*,. Washington, DC: Council of Graduate Schools.

summer/winter schools[203] and internships are logical training grounds, but some have been defunded (e.g., NSF Blue Waters Summer Internship), and they typically focus on the technical skills of early-career scientists.

Research experiences are a critical component of graduate school applications, yet access depends on institutional resources and faculty, which vary widely. Meanwhile, REU programs have become increasingly oversubscribed, necessitating selection criteria that balance previous research experience with how much an applicant has to gain from the opportunity.[204] More technical and research opportunities are needed for students from Primarily Undergraduate Institutions (PUI), Minority Serving Institutions (MSI; this includes Historically Black Colleges and Universities [HBCU], Hispanic Serving Institutions [his], and Tribal Colleges and Universities [TCU]), and Women's Colleges (WC), where many students study but fewer options exist, to increase their retention, graduation, and progression to graduate school and STEM careers.

Professional development is needed to keep astronomers current in the Profession's changing technical and career landscape. The current lack of "computational [training], knowledge, and access across the nation is a critical hindrance to the diversity and therefore the success of the field"[205] and a serious national security issue. Modern astrophysics demands computational literacy as a core competency, parallel in priority to math. Opportunities for observational training at modern facilities or developing technical skills to build state-of-the-art instrumentation are also limited. Training programs that address the planning, constructing, testing, and calibrating of new instruments are needed if complex projects are to be completed on time and at cost. More than 40 percent of astronomy Ph.D. recipients in 2015–2016 did not take postdoctoral positions, and many went into private sector jobs.[206] The Profession must respond to this trend and support a broad set of career pathways with an updated curriculum to include skills that are in demand.[207,208]

Goal 5, Suggestion 3: The panel suggests that the agencies fund PUI, MSI, and WC faculty and students in collaborations and research opportunities to access and engage in cutting-edge technological and data advancements, and that the Profession invest in expanded technical training pathways for all career levels.

Method, impact, and programmatics and cost to achieve this suggestion:

- **NSF-AST, NASA-APD/SMD, DOE-OoS**
  - *Method:* Create partnerships or training programs at MSI, PUI, and WC that facilitate long-lasting (5-year grants with administrative support) collaborations with major facilities (e.g., LSST, DESI), including National Laboratories (Fermilab or HPC facilities). Increase agency-funded REU programs and paid internships through partnerships with local industry (e.g., Metcalf program at the University of Chicago, TIMESTEP program at the University of Arizona).

---

[203] LSST Data Science Fellowship Program, LANL summer computational physics programs, Astro Hack weeks.

[204] McDevitt et al., 2020, *Ecology and Evolution*, 10(6):2710–2738.

[205] G. Besla, D. Huppenkothen, N. Lloyd-Ronning, E. Schneider, P. Behroozi, B. Burkhart, C.K. Chan, et al., 2019, Astro2020: Training the future generations of computational researchers, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1907.04460.

[206] P. Mulvey and J. Pold, "Astronomy Degree Recipients One Year After Degree" https://www.aip.org/statistics/reports/astronomy-degree-recipients-one-year-after-degree, accessed 26 August 2020.

[207] P. Heron and L. McNeil, 2016, "A Report by the Joint Task Force on Undergraduate Physics Programs," http://www.compadre.org/JTUPP/docs/J-Tupp_Report.pdf.

[208] APS/AAPT Effective Practices in Physics Programs, Career Preparation, https://ep3guide.org/guide-overview/career-preparation.

- — *Impact:* Access to computational resources leverages contributions of PUI-affiliates (faculty and students) and pan-STEM networks between PUI, MSI, and WC, and expands underrepresented students' opportunities to participate in STEM careers and graduate school.
- — *Programmatics:* $2 million–$3 million per collaboration;[209] $350,000 per year for expanded REU; and $500,000 per year to support five training programs.[210]

- **The Profession, Academic Institutions**
  - — *Method:* Expand technical training opportunities to include more than just early-career researchers. Embed computational training in standard curriculum with one first-year computer science course and one upper division course in computational methods, optimization, data science and/or statistics, with a focus on applications to physics and astronomy research.[211] Support faculty to develop new curricula through open-source platforms and communities of practice.
  - — *Impact:* Programming is recognized as a core competency, parallel in priority to math, and needed for all career levels.
  - — *Programmatics:* Low-cost. Work with computer science departments to facilitate implementation.

N.6.6 Goal 6: Cultivating Local and Global Partnerships

Reframe policies around community engagement in order to embed cultural humility, ethical practice, and a growth mindset throughout the Profession in a continuous effort to cultivate and sustain healthy cultures for scientific inquiry.

The demographics of the Profession reflect its values. Retention and participation of a professional community that comes from, interacts with, and returns to a diverse set of cultures can be achieved only by ensuring belonging for each of its members. Substantial, continuous effort is needed to enrich the culture of the Profession by ensuring that its members have the cultural fluency to advance values of both local communities and global needs. A reenvisioned model for engagement with communities at large, where ethical, sustainable, and healthy partnerships with local and global communities are held central, will ensure a more inclusive Profession and continued public support and trust.

N.6.6.1 Reenvision "Outreach" and "Broader Impacts" as Partnerships That Enable Growth and Enrichment Opportunities for the Profession

Astronomy is uniquely positioned in the public eye as both an awe-inspiring and a humbling science. As such, the Profession has made great efforts to engage the public through Education, Broader Impacts, and Outreach (EBIO) programs, many associated with major missions and collaborations.[212]

[209] Estimates based on MSI Partnerships in NSF-DMR PREM Program ($2–$3 million per collaboration).

[210] About seven new AST REU Sites awarded per year, $350,000/site for 3 years, 15 percent increase 350,000/year. Cost of training programs will depend on what the industry partner can provide in student salaries and administrative support.

[211] G. Besla, D. Huppenkothen, N. Lloyd-Ronning, E. Schneider, P. Behroozi, B. Burkhart, C.K. Chan, et al., 2019, Astro2020: Training the future generations of computational researchers, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1907.04460.

[212] NASA education programs (HST, Chandra, JWST planning). NSF EPO programming for NOIR Lab, NRAO, LSST EPO. DOE programming for LIGO, DESC. PI basis, NSF Broader Impacts programming is an agency-led directive to members of the profession to engage with communities.

Engagement with the public has generated public support, which translates into federal funding. Important strides have been made, making EBIO an excellent opportunity to connect more deeply with a broader community.[213] Including EBIO programming from conception stages of new astronomical endeavors, in partnership with EBIO professionals, would help integrate EBIO with the scientific team. Furthermore, EBIO programming could benefit from direction from the communities impacted in order to more effectively meet stakeholder needs.

However, there is currently an operating division between the dominant culture of the Profession, which reflects that of its predominant demographic groups, and cultural identities that are scarce within the Profession. Transforming the Profession into one that is multicultural, culturally fluent, and built on a partnership driven by equity-advancing values will enable equitable access and a sense of belonging for all. When individuals and communities are actively engaged ethically and with integrity, a "partnership" is established and community trust in science is strengthened.[214] Current EBIO programs and practices of engagement can be held to the ethical standards inherent to such partnerships.[215] Partnership recognizes that any person or community impacted by the Profession's programmatics or methods is a stakeholder, including participants and collaborators in gathering and contributing data, assessing need and impact, and making decisions. As such, the need for partnership is not limited to EBIO, but any practice by the Profession where humans and/or communities are impacted.[216] Partnership provides avenues for enrichment, self-reflection, and education for the Profession. The practice of partnership fundamentally requires learning about cultures and perspectives not well represented in either party's common experience through respectful engagement.

Goal 6, Suggestion 1: The panel suggests that the Profession reimagine community engagement and EBIO as partnerships. Partnerships are fundamental to the professional well-being of members and stakeholders, and provide the foundation from which the Profession could be transformed to be more inclusive, multicultural, and innovative. Effective partnerships rely on a foundation of oversight and accountability for the impact of EBIO activities on stakeholders, as outlined in Goal 1, Suggestion 2.

Method, impact, and programmatics and cost to achieve this suggestion:

---

[213] The current state of broader impacts, National Alliance for Broader Impacts January 2018 EBIO programs frequently have stated goals but are rarely held accountable to those goals; consequently, the impact of this programming is insufficiently evaluated. There are no procedures, guidelines, or means of assessment to ensure ethical treatment of those impacted by EBIO activities. See https://vcresearch.berkeley.edu/sites/default/files/wysiwyg/filemanager/BRDO/Current%20state%20of%20Broader%20Impacts%202018.pdf, accessed 24 August 2020.

[214] The NIH initiative All of Us (Precision Medicine Initiative) is a high-profile example of how the core values of partnership help build trust through transparency. See https://allofus.nih.gov, accessed 24 August 2020.

[215] Example manifestations of partnership are Community-Based Participatory Research (see Viswanathan, et al., 2004, "Community-Based Participatory Research: Assessing the Evidence: Summary," *AHRQ Evidence Report Summaries*, https://www.ncbi.nlm.nih.gov/books/NBK11852/, accessed 24 August 2020); Community Engaged Research (list of internal policies and external resources provided on *Community Engaged Research* by Ohio State University, Office of Responsible Research Practices, https://orrp.osu.edu/irb/research-participants/community-engaged-research/, accessed 24 August 2020); and community partnership in forestry practices (U.S. Endowment for Forestry and Communities, "The Status of Community Based Forestry in the United States," https://www.usendowment.org/the-status-of-community-based-forestry-in-the-united-states/, accessed 24 August 2020.

[216] Citizen science programs are an example of astronomy research that has benefited from community engagement. For example, "Astro 2020 State of the Profession White Paper: EPO Vision, Needs, and Opportunities Through Citizen Science" and "Astro 2020 Infrastructure Activity White Paper: Citizen Science as a Core Component of Research Infrastructure" by Laura Trouille (2020), which make use of GalaxyZoo, https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/, accessed 24 August 2020.

- **NASA, NSF, DOE, and Institutions**
  - *Method:* The panel suggests that EBIO[217] be reframed as a partnership. Reframing necessitates measurable benchmarks and outcomes that would be reported back to the agency in order to evaluate the establishment and effective operation of partnerships. Benchmarks, outcomes, or reports thereof would be based on identification of stakeholders and assessment of desired impact and stakeholders' needs in collaboration with stakeholders.[218] Outcomes would be included in established partnerships as part of evaluation for funding renewals and new proposals.
  - *Impact:* Ensures that ethical and mutually beneficial partnerships are established between the Profession and stakeholders.
  - *Programmatics:* Cost for the formation of agency specific office, as described in Goal 1, Suggestion 2.
- **NSF**
  - *Method:* Broader Impacts in its current form can be readily reframed as partnerships. Both Intellectual Merit and Broader Impacts need to be evaluated comprehensively and pass a threshold for the proposal acceptance.
  - *Impact:* All stakeholders are included in Broader Impacts programming. Consistent weighting and evaluation of Broader Impacts.
  - *Programmatics:* No-cost. Can be implemented in 1–2 years.
- **Research Facilities, Including Large Ground-Based Facilities and Space-Based Missions**
  - *Method:* The panel suggests that strategic planning for partnership programs (e.g., all EBIO efforts) begin at project/mission conception.[219] Further, partnerships ought to be properly staffed with personnel,[220] including EBIO professionals, and driven by a vision created by both the EBIO team and stakeholders. The panel suggests that stakeholder leaders of partnership programs be integrated into the project/mission leadership structure with access to the decision-making bodies, and that there be regular discussion of partnership outcomes with the scientific team. Proposal review criteria would reward evidence-based plans to establish such partnerships.
  - *Impact:* Partnerships are integrated within the operation of the Profession.
  - *Programmatics:* Benchmark of about 5 percent of the operational budget invested in building partnerships with stakeholder communities.[221]

---

[217] It is understood that "engagement" encompasses any "Outreach," "Broader Impacts," or other programming or projects where stakeholders can be identified.

[218] For example, one could envision guidelines for Broader Impacts sections that include the identification of stakeholders, evidence-based assessment of need, and impact of programming on addressing these needs. This framework would be relevant on all scales, from partnering with individuals at local schools, to collaborative programs between institutions or community organizations, to large facility construction.

[219] The Rubin Observatory's EPO efforts for the LSST are an effective example of integration of EPO programming at project conception. The VRO LSST EPO team contracted outside evaluators to conduct a user's needs assessment when designing their plans. EPO Design Document, Amanda Bauer https://docushare.lsstcorp.org/docushare/dsweb/Get/LEP-31/LEP-31.pdf, accessed 18 November 2020.

[220] Personnel needs may include technical staff (software engineers, web developers, IT staff—particularly in the era of massive data sets), education experts, social media expertise, and an assessment team that includes social scientists and community advisors.

[221] Large scientific projects with such a level of investment are expected to result in highly positive socioeconomic impact. M. Florio, 2019, *Investing in Science, Social Cost-Benefit Analysis of Research Infrastructures*, Cambridge, MA: MIT Press.

N.6.6.2 Build Inclusive Partnerships with International Stakeholders

Astronomy is a global, collaborative profession.[222] The U.S. community must attend to the well-being of international members and partnerships, or it puts ongoing scientific excellence at risk. Lack of cultural diversity and language barriers impede equitable inclusion in the Profession. In particular, institutions are accountable for ensuring the well-being of international participants by cultivating communities that do not erase cultural identity. Visa status and international political turmoil make international members vulnerable to abuse. Securing visas and associated documentation requires a significant time commitment and is an emotional stressor, thus impacting the mental health and research productivity of international members. During times of international distress, like the COVID-19 pandemic, consulates and international borders may be closed, disrupting visa applications.

> **Goal 6, Suggestion 2**: The panel suggests that the Profession promote global, culturally supported pathways into the Profession and provide training in inclusive community practice to all participants. The health of international collaborations could be enhanced by establishing and enforcing codes of conduct.

Method, impact, and programmatics and cost to achieve this suggestion:

- **The Profession and Funding Agencies**
  - *Method:* The panel suggests that funding agencies protect members of the Profession who are vulnerable owing to their international status. For example, institutions and professional societies[223] could identify teams of individuals who are versed in institutional resources for international participants to form Institutional Support Teams that aid international scholars and their families as they adjust to U.S. culture. The panel further suggests that U.S.-based and international partners could collaboratively agree upon codes of conduct, methods to enact them, and repercussions for violations.
  - *Impact:* A Profession that supports international participants to enable scientific excellence.
  - *Programmatics:* No-cost. Could be implemented immediately.
- **The Profession**
  - *Method:* The panel suggests building culturally supported training pathways for entry into the Profession for scholars from regions of the world where astronomy is growing, but engagement with the U.S. workforce is currently low (e.g., Chile, Mexico, South Africa). Healthy partnership programs would take into account the needs and cultural values of stakeholders. Examples include encouraging applications to graduate programs, developing pre- and postdoctoral exchange programs, and supporting summer schools and instructor training programs like PASEA.[224] Partnerships with the stakeholder communities are critical for developing culturally informed programming, including EBIO activities run at international observational facilities. Professional societies might create forums for outcomes and cultural knowledge to be exchanged with and reported to the Profession. The Profession also needs to encourage students and scientific leaders to be trained abroad, given the wealth of knowledge and expertise that is available.

---

[222] Many major collaborations have international partners, such as the LSST Dark Energy Survey Collaboration, SDSS, the International Space Station, and so on.

[223] AAS could partner with the IAU to make resources available to support international scholars—for example, through the USNC-IAU committee, https://aas.org/comms/usnc-iau-committee, accessed 13 November 2020.

[224] Pan-African School for Emerging Astronomers, https://www.astrowestafrica.org/about, accessed 24 August 2020.

- *Impact:* Increased participation from international communities within the Profession to advance scientific excellence.
- *Programmatics:* $150,000/year based on average IAU Astronomy Development Grants.[225]

N.6.6.3 Build Sustainable Partnerships with the Global Community and the Earth

Standard practices in the Profession have a negative impact on the environment and significantly contribute to climate change. Ironically, even as we search for habitable worlds, the Profession's large carbon footprint is decreasing the habitability of our own planet.[226]

The carbon footprint of the Profession is unnecessarily large owing to travel to conferences, meetings and observatories, high-performance computing, and facility construction.[227,228,229] The Profession has taken some steps to address this—for example, remote observing is becoming more common. COVID-19 forced a rapid response, in which the Profession has demonstrated responsible stewardship, notably through virtual conferences, panel reviews, and collaboration meetings (e.g., AAS 236th meeting). Remote conferences increase equitable access by removing constraints inherently associated with travel (e.g., dependent care, visa restrictions) and adding benefits (e.g., asynchronous participation, more affordable). These adjustments can be leveraged to build future models for responsible stewardship in the face of climate change. The Profession is often called the "gateway science" owing to the public's fascination with astronomy. This status affords the Profession a unique opportunity to educate the populace in scientific literacy, including climate change, thus fulfilling a major need in the current sociopolitical environment where there is a distrust of science that has real-life consequences.

Goal 6, Suggestion 3: The panel suggests that funding agencies, professional societies, and private foundations reduce the Profession's carbon footprint and other impacts on the environment. The panel suggests that the Profession increase engagement in initiatives that educate the public in the language of science with attention to climate change.

Method, impact, and programmatics and cost to achieve this suggestion:

- **Funding Agencies and Funding Institutions**
  - *Method:* The panel suggests that specific focus on reducing the carbon footprint be included in environmental assessments and mitigation plans in proposals for new facility construction, maintenance, and operation.[230] Further, facility operators and institutions

[225] Based on the approximate average for similar programming over 7 years for IAU Astronomy for Development Grants, https://www.iau.org/development/funding/, accessed 24 August 2020.

[226] Astronomers contribute more to climate change than the average global citizen, Climate Issue, 2020, *Nature Astronomy*, 4:811, doi: 10.1038/s41550-020-01216-9.

[227] L. Prichard, C. Oliveira, A. Aloisi, J. Roman-Duval, S. Hernandez, C. Pacifici, I. Momcheva, Enhancing Conference Participation to Bridge the Diversity Gap, white paper submitted to the Astro2020 Decadal Survey, https://arxiv.org/abs/1909.10996.

[228] Stevens, et al., 2020, The imperative to reduce carbon emissions in astronomy, *Perspective in Nature Astronomy*. Based on Australian astronomer emissions, but indicative of the carbon footprint of astronomers in general.

[229] S. Portegies Zwart, 2020, The ecological impact of high-performance computing in astrophysics, *Nature Astronomy*, 4:819–822.

[230] A summary of the benefits of combining Section 106 of the National Historic Preservation Act and the National Environmental Policy Act is given by the Advisory Council on Historic Preservation, https://www.achp.gov/integrating_nepa_106, accessed 16 May 2020.

participating in large collaborations could consider both a carbon offset plan and an assessment of the carbon footprint associated with travel.
  — *Impact:* Assessment, quantification, and mitigation of environmental impact by the Profession.
  — *Programmatics:* Costs can be included in a facility's planning process.
- **NSF-AST, NASA-APD/SMD, Educational/College/University Institutions**
  — *Method:* The panel suggests redesigning old and funding new initiatives and education programs to focus on climate change.[231]
  — *Impact:* Capitalize on public interest in astronomy to educate large audiences on scientific language and climate change.
  — *Programmatics:* No significant change in funding. Agencies would need to refocus priorities when assessing successful proposals.

N.6.7 Goal 7: Partnering with Indigenous Communities

Align the values of the Profession with those of Indigenous and other local communities impacted by the Profession to cultivate and sustain healthy partnerships for the benefit of both.

The future health of the Profession depends upon developing and maintaining healthy partnerships with Indigenous communities. Optimally sited observatories are a necessary resource for the Profession; access to those sites is critical to their success. Many ground-based observatories[232] are built on lands that have legal, cultural, historical, and/or sacred significance to Indigenous communities. Many large astronomy departments are hosted at academic institutions that have profited from similarly obtained land allotments.[233] Despite the value of these resources, Indigenous stakeholders are the least represented in the Profession,[234] suggesting that the Profession's past and current efforts to engage with Indigenous peoples are ineffective. Growing tensions owing to such land usage are recognized on a global scale, which negatively impacts public and political support for the Profession. It is therefore critical to develop long-term, targeted, functional partnerships with Indigenous communities that explicitly recognize Indigenous sovereignty and personhood.

Building healthy partnerships with Indigenous communities necessitates the following: (1) culturally supported pathways for inclusion in the Profession; (2) equitable access to education, current and emerging technologies, and economic benefits of hosting an astronomical facility; and (3) responsible stewardship in recognition of the use of Indigenous lands by non-Indigenous entities. This last includes partnership with Indigenous communities in order to make reparations and to enter respectful dialogue

---

[231] Although "any individual actions we take will pale in comparison to corporate and industrial pollution Astronomers have an 'ethical obligation … that must not be ignored' … we should not internalize environmental guilt; instead we must call for systemic change and fight against bad practice." Climate Issue, 2020, *Nature Astronomy*, 4:811, doi: 10.1038/s41550-020-01216-9.

[232] The following is a non-exhaustive list of some of the most prominent U.S. observatory sites listed with the associated Indigenous community: Maunakea Observatories (Kanaka Maoli); Kitt Peak National Observatory (Tohono O'odham); Mt. Graham International Observatory (Apache); Las Campanas Observatory, Cerro Pachón/Gemini South Observatory/Rubin Observatory (Diaguita); Atacama Large Millimeter Array (Likan Antai).

[233] For example, 10.7 million acres of Indigenous lands were allotted to 52 land grant universities through the Morrill Act and similar legislation to aid their economic development and growth. Institutions with astronomy programs that significantly benefit from these lands include Cornell, Penn State, Ohio State, Michigan State, Washington State, University of California, Rutgers, MIT, University of Maryland, University of Massachusetts, University of Wisconsin, University of Arizona, and University of Minnesota.

[234] There are approximately 6.8 million Indigenous people (U.S. Census Bureau, 2020) living in the United States (~2.09 percent total population), but on order of 10 hold Ph.D.s in physics and astronomy, https://worldpopulationreview.com/states/native-american-population/, accessed August 2020.

about the construction of future facilities. The above provides a foundation upon which community values may be realized.

N.6.7.1 Mitigate the Negative Impact of Past Engagement Around the Summit of Maunakea as Part of a Larger Effort to Build a Functional Partnership with Local Indigenous Communities

Lack of an authentic partnership with Kanaka Maoli (the Indigenous people of Hawai'i) impedes the efficacy of the astronomy workforce, significantly risks facilities' investments, negatively impacts Kanaka Maoli, and diminishes public support. It puts into question the integrity upon which scientific discovery is realized. The Profession has not practiced responsible stewardship as described by the equity-advancing values proposed in this report. This is manifested by the lack of guiding principles, to be upheld by the University of Hawai'i (UH), the Thirty Meter Telescope (TMT) International Observatory (TIO), or the Profession as a whole, for the ethical practice of astronomy. Box N.3 gives a brief summary of the Profession's activities on Maunakea in the historical context of engagement with Kanaka Maoli. It highlights the negative impact past modes of engagement have had on both the Profession and Kanaka Maoli, with the intent to learn from past mistakes and frame a pathway for a more equitable and collaborative future together for the benefit of all.

BOX N.3 Contextual History of Engagement of the Profession with Kanaka Maoli

The summit of Maunakea has become home to 13 of the most successful observatories in the world. The anticipated addition of the Thirty Meter Telescope (TMT) is expected to revolutionize ground-based astronomy and was listed as a top priority in the Astro2000 report.[1] However, ongoing demonstrations by kia'i (guardian or protector), led by Kanaka Maoli, bring a serious concern into focus. The construction of TMT, the means by which it is realized, and its impact on Kanaka Maoli have to be recognized within the larger context of Hawaiian history.[2,3,4] Failure to do so[5,6,7] has led to several iterations of major delays,[8,9] arrests,[10,11,12] rulings,[13,14,15] and governor-issued directives[16] since the groundbreaking ceremony was disrupted[17] in 2014.[18] New construction on Maunakea has seen falling public support.[19] As of December 2019, Governor Ige of Hawai'i has withdrawn state law enforcement owing to the $15 million cost surrounding safe access to the summit[20] simultaneously met by halted construction of TMT.[21] Furthermore, continued access to observatories in current operation at the summit is intermittently obstructed.[22] All investments to date are at risk if these issues are not resolved with a long-term plan in place. Instead, the value of these investments and the integrity of the Profession is realized should the Profession work in collaboration with Kanaka Maoli.

The construction of TMT falls during a time of Indigenous cultural reclamation in Hawai'i after over a century of persecution and systemic oppression. The summit of Maunakea[23] is traditionally regarded by many Kanaka Maoli as sacred,[24,25,26,27] as "a place for gods and not for construction of edifices for human use."[28] Ahu (shrines), heiau (temples), pu'u (hills), and burial sites around the summit are a testament to the reverence held by Kanaka Maoli for the Mauna. Cultural practices, like burial of 'iewe (placenta) and gathering of sacred waters, require continued, free access, which is inhibited by current policies around the summit. The Profession's impact on Kanaka Maoli culture is substantial.

A narrative constructed from TMT budgets suggests that the full impact of the project on Maunakea and Kanaka Maoli has been systematically minimized and ignored. Based on documents provided by the TMT to this panel, the preconstruction planning and development phase alone totaled $211.1 million, with clearing and building costs between 2014–2020 totaling $19.3 million. During this same time, a relatively small investment ($13.3 million[29]) was devoted to community engagement efforts. Moving forward, the expected annual cost for operations and maintenance of TMT is $47.0 million plus a sublet cost of $1 million/year to be paid to UH. The lease agreement for the observatory complex on Maunakea

between the state and UH is $1/year ending in 2033, when all lands shall be returned to original conditions within reason.[30] The proposed construction of TMT, just before the lease termination date, sends a message of devaluation to Kanaka Maoli. Every legal effort and counterclaims filed by Kanaka Maoli, including pressures to decommission an observatory before new construction, appeals to reopen construction, and any other disagreements regarding construction on Maunakea, are reported to have "minimal" impact on observatory budgets. Further, no cost estimate has been made for post-lifetime TMT life-support, suggesting that there has been little consideration for the long-term stewardship for Maunakea. When value is equated with dollar signs, the value placed by the Profession on Kanaka Maoli culture, values, voices, and needs is "minimal" except under threat of discontinued operation and construction of observatories on Maunakea.

---

[1] NRC (National Research Council), 2001, *Astronomy and Astrophysics in the New Millennium*, Washington, DC: The National Academies Press.

[2] R. Alegado, 2019, Opponents of the Thirty Meter Telescope fight the process, not science, *Nature*, 572, https://www.nature.com/articles/d41586-019-02304-1, accessed 24 August 2020.

[3] A. Witze, How the fight over a Hawaii mega-telescope could change astronomy, *Nature*, 577, https://www.nature.com/articles/d41586-020-00076-7, accessed 24 August 2020.

[4] C. Prescod-Weinstein, et al., 2020, "Reframing Astronomical Research Through an Anticolonial Lens—For TMT and Beyond," Community input from submission, 27 January.

[5] B. Isaki, S. Muneoka, and K.H. Kanahele, 2020, "Kū Kia'i Mauna: Historical and Ongoing Resistance to Industrial Astronomy Development on Mauna Kea, Hawai'i," Community input from submission, 8 January.

[6] K. Kiyuna, 2020, "Ka Piko Kaulana o ka 'Āina: Additional Context for Understanding the Cultural Significance of Mauna Kea," Community input from submission, 8 January.

[7] S. Kahanamoku, R.A. Alegado, K.L. Kamelamela, B. Kamai, L.M. Walkowicz, C. Prescod-Weinstein, M.A. de los Reyes, and H. Neilson, 2020, "A Native Hawaiian-Led Summary of the Current Impact of Constructing the Thirty Meter Telescope on Maunakea," Community input from submission, 9 January. 64f0d1cc4f85beae7842d196a156c767_Native_Hawaiian_Impacts_Astro2020_final.pdf, accessed 24 August 2020.

[8] TMT construction delayed—11 Apr 2015, *khon2*, http://khon2.com/2015/04/11/thirty-meter-telescope-construction-delayed/, accessed 24 August 2020.

[9] TMT construction delayed—19 Dec 2019, Hawai'i Public Radio, https://www.hawaiipublicradio.org/post/tmt-wont-begin-construction-time-protesters-told-clear-mauna-kea#stream/0, accessed 24 Aug 2020.

[10] C. Jones, Associated Press, 2015, "Clash in Hawaii Between Science and Sacred Land," 3 April, in *US News*, https://www.usnews.com/news/science/news/articles/2015/04/03/clash-over-telescope-at-sacred-hawaiian-site-intensifies, accessed 24 August 2020.

[11] "Police, TMT Issue Statements on Mass Arrests on Mauna Kea," *Big Island Video News*, 2 April 2015, http://www.bigislandvideonews.com/2015/04/02/police-tmt-issue-statements-on-mass-arrests-on-mauna-kea/, accessed 24 August 2020.

[12] "Department of Land and Natural Resources Releases Names of Those Arrested on Maunakea," *Hawaii Tribune Herald*, 24 July 2019, https://www.hawaiitribune-herald.com/2019/07/24/hawaii-news/dlnr-releases-names-of-those-arrested-on-maunakea/, accessed 24 August 2020.

[13] See Hawai'i Board of Land and Natural Resources Case BLNR-CC-16-002.

[14] Hawai'i Supreme Court case SCAP-14-0000873, https://www.courts.state.hi.us/docs/opin_ord/sct/2015/December/SCAP-14-0000873.pdf, accessed 25 August 2020.

[15] Hawai'i Supreme Court case SCOT-17-0000777, https://law.justia.com/cases/hawaii/supreme-court/2018/scot-17-0000777.html, accessed 25 August 2020.

[16] Governor Y. Ige of the State of Hawai'i, State of the State Address, "News Release: Governor David Ige Announces Major Changes in the Stewardship of Mauna Kea," https://governor.hawaii.gov/newsroom/news-release-governor-david-ige-announces-major-changes-in-the-stewardship-of-mauna-kea/, accessed 24 August 2020.

[17] "TMT Groundbreaking Disrupted," *Hawaii Tribune Herald*, 8 October 2014, https://www.hawaiitribune-herald.com/2014/10/08/hawaii-news/tmt-groundbreaking-disrupted/, accessed 24 August 2020.

[18] References listed here on the history of negotiations on the construction of TMT on Maunakea are not exhaustive.

[19] K. Dayton, "Public Support for TMT Drops Sharply, According to New Honolulu Star-Advertiser Poll," *Star Advertiser*, 25 September 2019, https://www.staradvertiser.com/2019/09/25/hawaii-news/public-support-for-tmt-

drops-sharply-according-to-a-new-honolulu-star-advertiser-poll/?HSA=78e87324d7c8c9011961208b9ed13d2797888c07.

[20] Associated Press, "Governor Says Hawaii Spent $15M to Ensure Mauna Kea Access," *Hawai'i Public Radio*, 18 December 2019, https://www.hawaiipublicradio.org/post/governor-says-hawaii-spent-15m-ensure-mauna-kea-access#stream/0, accessed 25 August 2020.

[21] C. Harlow and K. Hiraishi, "TMT Won't Begin Construction at This Time, Protestors Told to Clear Mauna Kea," *Hawaii Public Radio*, 19 December 2019, https://www.hawaiipublicradio.org/post/tmt-wont-begin-construction-time-protesters-told-clear-mauna-kea#stream/0, accessed 25 August 2020.

[22] *Nature* 577, 457–458, 2020.

[23] These lands were deemed "public" following the coup overthrow of the Hawaiian sovereign nation in 1893 and reassigned as "Ceded Lands" when Hawai'i became a U.S. territory in 1959.

[24] B. Isaki, S. Muneoka, and K.H. Kanahele, 2020, "Kū Kia'i Mauna: Historical and Ongoing Resistance to Industrial Astronomy Development on Mauna Kea, Hawai'i," Community input from submission, 8 January 2020. .

[25] K. Kiyuna, "Ka Piko Kaulana o ka ʻĀina: Additional Context for Understanding the Cultural Significance of Mauna Kea," Community input from submission, 8 January 2020

[26] T.K.H. Kanahele and D. McGregor, "Impacts of Astronomy on Indigenous Customary and Traditional Practices as Evident at Mauna Kea," 6 January 2020, https://doi.org/10.6084/m9.figshare.11522289.v1, accessed 26 August 2020.

[27] S. Kahanamoku, R.A. Alegado, K.L. Kamelamela, B. Kamai, L.M. Walkowicz, C. Prescod-Weinstein, M.A. de los Reyes, and H. Neilson, 2020, "A Native Hawaiian-Led Summary of the Current Impact of Constructing the Thirty Meter Telescope on Maunakea," Community input from submission, 9 January..

[28] This relationship has been documented as early as 1826 by missionary Joseph Goodrich. B. Isaki, S. Muneoka, and K.H. Kanahele, 2020, "Kū Kia'i Mauna: Historical and Ongoing Resistance to Industrial Astronomy Development on Mauna Kea, Hawai'i," Community input from submission, 8 January 2020..

[29] $5.5 million went toward the Community Benefits Package and $7.8 million went toward education and public engagement.

[30] State of Hawai'i General Lease No. S-1491.

The misalignment between the Profession's actions and Indigenous values has led to the current impasse. The situation on Maunakea in Hawai'i jeopardizes the following: (1) *Economic prosperity* through the potential loss of all investment in future observatories and access to current observatories. (2) *Health and well-being* of Indigenous astronomers who are forced to choose from a false dichotomy between cultural and professional values, thus creating both an unnecessary conflict within Indigenous communities and a narrative that counters any efforts toward inclusion of Indigenous people,[235,236] the least represented group within the Profession. Members of the Profession are forced to align for or against construction of TMT, which can be divisive within the scientific community when moral principles are not in alignment with science driven goals. (3) *Broadening participation and continued innovation* because both the academic pursuit of excellent science and Indigenous practices are lost or impeded by ongoing conflict around access[237] to the lands on and around Maunakea's summit.

The following methods suggest a path forward that begins and ends with Indigenous stakeholders and protectors of the land. It relies upon the inherent integrity of the Profession to pause all construction, listen to Indigenous communities, and engage in ethical practices that build trust and fundamentally acknowledge Indigenous personhood. These methods are meant to serve as a foundation upon which

[235] H. Kaluna, M. Neal, M. Silva, and T. Trent, 2020 "A Collective Insight into the Cultural and Academic Journeys of Native Hawaiians While Pursuing Careers in Physics and Astronomy," Community input form submission,6 March.

[236] A. Venkatesan, D. Begay, A. Burgasser, I. Hawkins, K. Kimura, N. Maryboy, L. Peticolas et al., "Collaboration with Integrity: Indigenous Knowledge in 21st Century Astronomy," white paper submitted to the Astro2020 Decadal Survey, https://baas.aas.org/pub/2020n7i020/release/1.

[237] Cultural practices that require access to the summit and its surrounding lands can be unplanned and personal in nature, requiring unfettered and timely access.

future and current facilities, institutions, observatories, and observatory sites can assess investments in partnership with Indigenous stakeholders.

Goal 7, Suggestion 1: The panel suggests that funding agencies hold ground-based observatories accountable to a high ethical standard, particularly around the construction of TMT on Maunakea. A true partnership as defined above would redirect effort to identify stakeholders and assess their needs, values, and activities, especially in relation to the Kanaka Maoli.[238]

Method, impact, and programmatics and cost to achieve this suggestion:

- **State of Hawai'i and TMT Institutions, Held Accountable by Funding Agencies**
  - *Method:* The panel strongly suggests that *any* new or continued construction on the summit of Maunakea be contingent upon having proactively established a pathway forward using a community-based approach that is based on consent and *mutual* agreement.[239] To ensure said pathway, the panel suggests, in addition to following guidelines developed in Goal 1, Suggestion 2, and Goal 6, Suggestion 1, the three methods outlined below. The panel further suggests that funding agencies not invest in future projects on Maunakea unless this and the following three methods are realized.
  - *Impact:* Allow time for respectful dialogue, which cannot occur under duress.[240]
  - *Programmatics:* No change in cost.[241]
- **TMT International Observatory LLC (TIO), University of Hawai'i (UH), and other Facility Lease Holders on Maunakea's Summit, Held Accountable by Funding Agencies**
  - *Method:* Allocate funding in facilities budget for proactive, ecologically sound maintenance of current facilities and complete cleanup of decommissioned observatory sites.[242] The panel suggests that funding agencies mandate annual reports on maintenance, cleanup, and other terms of land lease/occupation, as a requirement of any federal investment in TMT and in compliance with Goal 6, Suggestion 3.
  - *Impact:* Demonstrate that Indigenous voices have been heard on this matter and are respected, and thus intentional reparations are enacted.
  - *Programmatics:* Federal agencies can ensure compliance. Cost is $1 million/year for maintenance, $23.5 million/observatory for decommissioning and cleanup.[243] These costs will need to be verified and updated using independent estimates and in collaboration with the local community.

---

[238] The NSF statement on August 13 (https://www.nsf.gov/news/news_summ.jsp?cntn_id=301034) is an encouraging motion in the proposed direction with the hope that these efforts will be used to effectively engage with local Indigenous stakeholders and define a mutually beneficial pathway forward. Should a formal federal environmental review process begin, inclusion of local Indigenous stakeholder perspectives is critical for assessing outcomes and process.

[239] There have been proposals, such as those of Governor Ige in 2015 (see Box N.3), in reaction to demonstrations. The panel's overarching suggestion is that the profession position itself to proactively approach the coming decade, rather than continue down a trajectory that is increasingly reactive in nature.

[240] See Box N.3 for a brief historical accounting, and references therein that were provided by Kanaka Maoli to this panel, as evidence of Indigenous perspectives and experiences.

[241] TMT declined to report delay costs in the report they provided to the panel. It is here assumed that these will not exceed current costs.

[242] A. Witze, 2015, Hawaii prunes Mauna Kea telescope hub, *Nature*, 522:15–16, https://www.nature.com/news/hawaii-prunes-mauna-kea-telescope-hub-1.17688, accessed 26 August 2020.

[243] One to three of the 13 current observatories on the summit are projected to be decommissioned in the next few years, whereas the current lease mandates all 13 to be completely cleaned up by 2033. This cost is based on the estimate provided by TIO for a single observatory. The expected cost investments for maintenance, decommissioning, and cleanup were provided by TMT and are in 2019 dollars.

- **TIO,** Held Accountable by **Funding Agencies**
    - — *Method:* Fund initiative(s) for stakeholders who have an interest in Maunakea, including Kanaka Maoli cultural knowledge holders, to open a respectful and continuous dialogue around informed consent, where Kanaka Maoli are included in the TMT/TIO leadership. Informed consent[244] means an iterative process of proposal and review that addresses ethics and impacts on Indigenous persons and communities. Funding agencies can hold TIO accountable by making any federal funding for TMT contingent upon the ethical practices for partnership.
    - — *Impact:* Provide a roadmap for the respectful development of future facilities that upholds the integrity of Indigenous people and the Profession.
    - — *Programmatics:* Cost: $10 million initial efforts, 10 percent annual operating and maintenance costs—in addition to “Community Benefits Package” and “Education and Public Outreach.”
- **Funding Agencies and Institutions**
    - — *Method:* Systematically determine whether there are Indigenous stakeholders and what their needs, values, and activities are prior to and during development of any new facility. Hold facility development to the same ethical standards as any partnership in the Profession.[245] Within this framework, local stakeholders (especially Indigenous) would be included in planning, construction, maintenance, and decommissioning of facilities, as well as in defining benchmarks for accountability.
    - — *Impact:* Funding Institutions and land holders would create an ethics review board, in accordance with Goal 6, Suggestion 1, tasked with review and approval of facilities development, working in partnership with local stakeholders. Funding agencies can provide federally mandated and professionally established ethical standards, protections, and guidelines for individual human, cultural, artifact, and environmental impacts from facilities development.
    - — *Programmatics:* Included in construction and maintenance cost.
- **The Profession and Funding Agencies**
    - — *Method:* Require proposals using observational facilities that have Indigenous stakeholders consider the societal impacts of the observatory and its use on those communities. The panel suggests that a mandatory educational module be included in the time application, where this module would be developed in collaboration with Kanaka Maoli and focus on societal impacts and the equity-advancing values outlined in the section “A Values Statement for the Profession of Astronomy and Astrophysics,” earlier.
    - — *Impact:* Self-education of PIs on the process and impact of observatory construction on Indigenous lands.
    - — *Programmatics:* Low-cost. Could be implemented immediately.

N.6.7.2 Build Functional Partnerships with Indigenous Communities and Culturally Supported Pathways for the Inclusion for Indigenous Members of the Profession

The panel believes that there is a critical need to build long-term, functional partnerships with Indigenous communities. Lack of resources, often related to the limited availability of culturally relevant

[244] Defined in the Department for Health and Human Services Common Rule Federal Policy for the Protection of Human Subjects.

[245] Examples are literature surveys, stakeholder surveys, focus groups, cultural impact surveys like those required by the National Historic Preservation Act, and an evaluating committee that includes historians, environmental protection representatives, local community representatives, and sociologists.

education systems, as well as poverty[246] are major contributing factors to the broad education gap in Indigenous communities starting in early childhood.[247] Tribal Colleges and Universities (TCU) and Indigenous education centers are increasingly able to provide education through culturally relevant systems of study, but these same institutions are often severely underresourced. For example, such institutions may not have basic Internet access, technological infrastructure, or support for adequate computational literacy education and training. Students from underresourced institutions suffer the consequences of inequitable access from the earliest career stages. The combination of underresourced educational institutions and cultural marginalization within the Profession ultimately counters the inclusivity efforts of Indigenous scientists within the Profession. The addition of the optics of world-class facilities, occupied by non-Indigenous people, on Indigenous lands, can deepen distrust for the Profession in some Indigenous communities. Initiatives that aim to build mutually respectful and culturally relevant partnerships with Indigenous communities are shown[248] to significantly increase support for the Profession from local Indigenous stakeholders—and more broadly STEM—and to open culturally supported pathways for Indigenous youth to enter the Profession.

Goal 7, Suggestion 2: The Profession is accountable for promoting equitable, culturally supported participation. This requires a change in the Profession's culture so that Indigenous contributions are appropriately credited and Indigenous people and their cultures and values are granted respect. The panel suggests that funding agencies increase the scope of engagement and funding for existing partnerships with Indigenous communities and new partnership initiatives. Indigenous participation can be supported using targeted funding for (1) fellowships that support astronomy students from Indigenous communities, (2) Indigenous-led research, and (3) partnerships and support networks between Indigenous educational centers and larger research institutions and collaborations.

Method, impact, and programmatics and cost to achieve this suggestion:

- **The Profession, AAS Journals**
  - *Method:* Self-educate about Indigenous methods of producing, curating, and sharing Indigenous Traditional Knowledge (TK), which include oral histories and protocols, in order to develop, in partnership with Indigenous communities,[249] standards for respectfully crediting and using TK (e.g., in journal articles and talks).[250] The panel suggests that the Profession change language that reinforces adversarial or dismissive attitudes toward Indigenous communities and perspectives.

[246] Indigenous communities experience more than twice the national poverty rate. United States Census Bureau, https://data.census.gov/cedsci/table?q=B17&d=ACS%201-Year%20Estimates%20Detailed%20Tables&tid=ACSDT1Y2018.B17001C&vintage=2018, accessed 24 August 2020.

[247] UN Department of Economic and Social Affairs, Indigenous Peoples, Education Report, https://www.un.org/development/desa/indigenouspeoples/mandated-areas1/education.html, accessed 24 August 2020.

[248] Lee et al., 2020, "Building a Framework for Indigenous Astronomy Collaboration: Native Skywatchers, Indigenous Scientific Knowledge Systems, and The Bell Museum," International Planetarium Society Conference Proceedings. [[SRO: PLEASE PROVIDE FULL CITATION]]

[249] For example, use Traditional Knowledge Labels, https://localcontexts.org/tk-labels/, accessed 24 August 2020.

[250] Standards and protocols set forth by the Global Indigenous Data Alliance through the FAIR and CARE principles are an emerging avenue for such endeavors, https://www.gida-global.org, accessed 24 August 2020.

  - *Impact:* Lay groundwork to meaningfully and respectfully credit culturally significant Indigenous contributions.[251]
  - *Programmatics:* No cost. Can be implemented immediately.
- **NSF and NASA**
  - *Method:* Fund PIs located at TCUs, from Indigenous communities, or at institutions that predominantly serve Indigenous populations in partnership with Indigenous communities in order to develop culturally supported, Indigenous-led research and extended (5 year) research engagement through faculty and student training and mentorship,[252] administrative support, and up-to-date technological tools and support. Optimally, these would include funding for efforts to build strong, long-term research partnerships with large institutions/big data centers/collaborations with the aim of developing culturally supported pathways for full participation of Indigenous people in science careers.
  - *Impact:* Provide equitable access and increase multimodal expertise.
  - *Programmatics:* $200,000/year to support two initiatives at $100,000/year per agency and implemented in 2–3 years.
- **NSF, NASA, and DOE**
  - *Method:* Fund Indigenous education centers in partnership with Indigenous communities. This could include building and maintaining a computational infrastructure to enable remote participation in education opportunities, conferences, collaboration, and training from within Indigenous communities.[253] This includes computational facilities, AV equipment, and training, with Internet standards of an R1 institution.
  - *Impact:* Provide equitable access and amplify Indigenous voices and approaches within the Profession.
  - *Programmatics:* Cost: $1 million/year per agency, implemented in 1–2 years.[254]
- **NSF, NASA, and DOE, Private Foundations**
  - *Method:* The panel suggests that private foundations create long-term, $50,000/year fellowships, from undergraduate to Ph.D., for students belonging to Indigenous communities. The panel further suggests that federal agencies create bridge fellowships for students from TCU and Tribal Community Centers.
  - *Impact:* Provide equitable access to and amplify Indigenous voices and approaches within the Profession.
  - *Programmatics:* $100,000/year per agency and $500,000/year from private foundations, implemented in 1–2 years.[255]

[251] A. Venkatesan, D. Begay, A. Burgasser, I. Hawkins, K. Kimura, N. Maryboy, L. Peticolas. et al., ., “Collaboration with Integrity: Indigenous Knowledge in 21st Century Astronomy,” white paper submitted to the Astro2020 Decadal Survey, https://baas.aas.org/pub/2020n7i020/release/1.

[252] A. Venkatesan, D. Begay, A. Burgasser, I. Hawkins, K. Kimura, N. Maryboy, L. Peticolas et al., “Collaboration with Integrity: Indigenous Knowledge in 21st Century Astronomy,” white paper submitted to the Astro2020 Decadal Survey, https://baas.aas.org/pub/2020n7i020/release/1..

[253] Many Indigenous cultures value physical presence within their home community. In these cases, equitable access can only be attained when this cultural value is supported via remote participation.

[254] Grants would provide financial support for infrastructure and maintenance. This program is designed to equip and support all TCUs over a decade, with institutional needs widely varying. Costs have been calculated on the basis of 37 institutions, each of them being provided a total of about $500,000 over a decade.

[255] Grants would provide financial support for 18 students per year in physics and astronomy.

N.7 SUMMARY, BENCHMARKS, AND CONCLUSION

The Profession needs to establish workplaces and infrastructures that reflect equity-advancing values to allow the full human diversity of the nation to meaningfully contribute to the field in the interest of maximizing technical innovation and scientific excellence.

The panel has outlined a multi-faceted program to *leverage funding* to recognize, motivate, support, and hold accountable workplaces built on equity-advancing values; *reimagine leadership* to benefit from the multimodal expertise of our full community; use that leadership to *end discrimination and harassment*; *remove barriers to education and training* to ensure equitable access to knowledge and full participation in astronomy at all career stages; *build meaningful partnerships with astronomy's local and global communities* in recognition that a truly inclusive astronomy is inseparable from every one of the spheres it inhabits; and *partner with Indigenous communities* in order to cultivate and sustain healthy partnerships (e.g., Goal 6) for mutual benefit.

Ultimately, *data* is needed to inform every stage of this program; it has to be collected routinely, comprehensively, and with intention. The success of this more robust engagement with data, the panel suggests, depends on a *dedicated office*, in each agency, to oversee implementation and use the data repository to monitor progress toward realizing the goals.

N.7.1 Suggested Timeline and Benchmarks

The relationship to the goals is outlined in Table N.1.

- *Year 1:* (1) Set expectations for scientific conduct; (2) implement moderate, low-cost changes; and (3) assemble the resources and structures to plan for and support change.
- *Years 1–3:* (1) Adopt comprehensive program requirements; (2) rebalance funding priorities to expand prior and begin new programs; and (3) apply resources to support, review, assess, and hold accountable.
- *Mid-decade:* The panel suggests that dedicated agencies independently and in collaboration organize advisory board groups that can work with a National Academies-appointed mid-decadal panel to assess the progress and compare with initial benchmarks; preferably as a publicly available report to the advisory board groups. Findings would be used to update existing plans and inform directions for years 6–10.

N.7.2 Conclusion

The pursuit of science, and by extension scientific excellence, is inseparable from the humans who animate it.

This recognition, as stated in the introduction to this panel's report, motivates the suggested work to systematically embed equity-advancing values throughout astronomical research, technical, and education programs. The necessary growth and change to reframe existing structures will not always be comfortable. However, astronomers have always asked big questions and pursued fundamental challenges. The goals stated here are no less worthy of our vigorous intellectual engagement and commitment than any of the other daunting problems we pursue, from the origins of life to the nature of dark matter and dark energy. Only by properly supporting the people who do the science can we maximize the return on the nation's investment in fundamental research.

TABLE N.1 Timeline of Actions That Require Structural Changes

Year	Area	Essential Actions (Goals)
1	Set expectations for scientific conduct.	Recognize identity-based harassment as scientific misconduct and build the necessary structures to hold individuals and institutions responsible for harassment and discrimination (Section N.6.4). Make any construction on Maunakea contingent on adopting a community-based approach, including ecological considerations (Section N.6.7).
	Immediate, low-cost items.	As listed in the document and associated tables (all goals, sections N.6.1–N.6.7).
	Assemble resources for change, providing support, review, assessment, and accountability.	Establish a dedicated office to collect demographic information for each agency (DOE, NASA, and NSF) following NIH's framework. Initiate systematic data collection and storage, and create baselines by aggregating existing data nationwide (Section N.6.1). Build a working group that includes professional societies and private foundations in close collaboration with the National Academies to assess how support structures can be created for targets of discrimination/harassment (Section N.6.4). Agencies and institutions of the Profession should engage with experts and community stakeholders to work toward creating equity-advancing programmatics including Partnerships (sections N.6.1, N.6.2, N.6.4, N.6.6, N.6.7).
1–3	Comprehensive program requirements.	Adopt requirements for individuals, teams, facilities, and institutions to address equity-advancing values, including Partnerships with relevant stakeholders, in their proposals, funded activities, and award reports (sections N.6.2, N.6.3, N.6.6).
	Rebalance funding priorities.	Institute Training Grants, Early Career Awards, Leadership Programs, Physics Education Research, PAARE/VFP, REU, Partnerships with PUI/MSI/SC/TCU and relevant Indigenous, local, and global communities (sections N.6.2, N.6.3, N.6.5, N.6.7).
	Apply resources to support, review, assess, and hold accountable.	Ensure that all funded research is conducted in accessible spaces and that reporting and assessment of mentoring is built into proposals and reporting systems (Section N.6.1). Provide mechanisms for data-driven accountability to ensure that programmatics reflect equity-advancing values as derived from agency founding documents (sections N.6.2, N.6.4). Allocate resources in Partnership with Indigenous, local, and global communities (sections N.6.6, N.6.7).
5, 10	Review progress.	The panel suggests that agencies independently and in collaboration organize advisory board groups (Section N.6.1.2) that can work with a National Academies-appointed mid-decadal panel to assess the progress on the various components of these programs and aggregate data nationwide and longitudinally and compare with initial benchmarks (all goals, sections N.6.1–N.6.7), preferably as a publicly available report to the advisory board groups (Section N.6.1.2).
6–10, 11+	Apply lessons learned from mid-decadal review to adjust actions to achieve goals.	Establish diverse leaders who practice equity-forward values; increase inclusion and persistence of scientists and scholars from historically underrepresented groups; provide continual technical training for all members of the Profession; reenvision "outreach" and continually reframe community engagement as partnerships with the Profession (all goals, sections N.6.1–N.6.7).

O

Independent Technical, Risk, and Cost Evaluation

The Astro2020 decadal survey evaluated the technical, programmatic and execution risks of large ground and space-based projects in numerous ways. This evaluation was directed by the relevant Program Panel, and included multiple layers of review by the panel, as well as an Independent Technical, Risk and Cost Evaluation (TRACE), performed by a contractor (The Aerospace Corporation). This full level of independent review was undertaken for all National Science Foundation (NSF) Major Research Equipment and Facilities Construction (MREFC)-scale projects selected by the Program Panels to be of primary interest, as well as for some version of all of the NASA flagship concepts (in the case of the Large UV/Optical/IR Surveyor (LUVOIR) and Habitable Exoplanet Observatory (HabEx), multiple implementations were studied, and the EOS-1 program panel selected those of primary interest). For the NASA flagships, in addition to the project cost estimates, NASA performed a Large Mission Concept Independent Assessment Team (LCIT) evaluation, independent of the projects, which, in addition to the project estimates, was considered as input to the panels. The panels included subject-matter experts in technology, management, and instrumentation, and this expertise guided the process, providing key feedback and iterative steps between both the projects and the independent contractor. Information flow is captured in Figure O.1. The Program Panel process is documented in the panel reports and summarized in the main text of the report. The purpose of this appendix is to describe the process through which the Aerospace Corporation's independent TRACE evaluations were implemented and how they were managed and considered by the survey.

The TRACE provided broad guidance to the Program Panels in the form of a forward-looking assessment of potential technical risks and likely cost and schedule boxes associated with a mission concept, but not definitive cost nor schedule estimates.

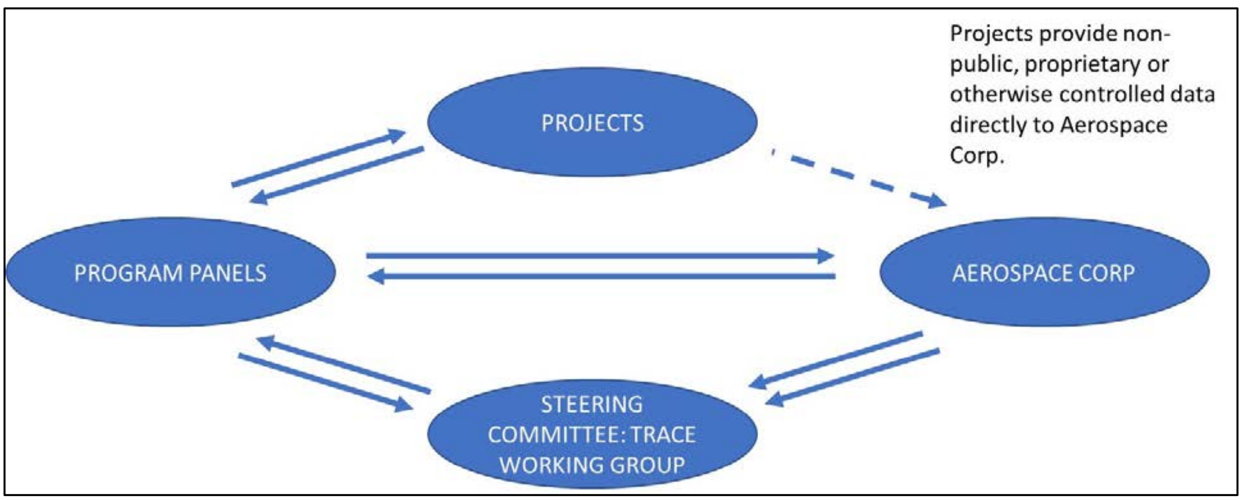


FIGURE O.1 Relationship between projects, program panels, aerospace, and the steering committee.

The decadal independent cost and technical risk assessment process has substantially evolved and improved since its earliest application, as well as its implementation for the *New Worlds, New Horizons* (NWNH) 2010 astrophysics decadal survey. The current methodology, TRACE, is anchored by successful implementations of its heritage risk analysis methodology, the Cost and Technical Evaluation (CATE) process, for a total of four decadal surveys, including NWNH. It is also anchored in the processes developed in support of the independent estimating of both NASA and Department of Defense projects. The details of the Aerospace Corporation process have been captured in multiple reports and presentations, including *The Space Science Decadal Surveys-Lessons Learned and Best Practices*[1] and *CATE Overview, Astro2020 Large Mission Concepts*.[2]

Highlights of the TRACE process included using Aerospace Corporation proprietary models and databases, validated through multiple processes, including comparisons to historical implementations, discussed below. Threats and risks were evaluated using Monte Carlo simulations. Each project was assessed independently, regardless of maturity, and brought to a baseline 70 percent probability of success through application of reserves for both cost and schedule to assure equal programmatic evaluation criteria could be applied.

In addition, NASA's support of the development of high-quality mission concepts significantly assisted the TRACE. The Astrophysics Science and Technology Definition Teams (STDT) in preparation for this decadal survey provided well-developed, compelling, and executable mission concepts with high-fidelity estimates.

The Astro2020 decadal survey committee used best practices from NWNH and implemented lessons learned from other completed decadal surveys.[3,4]

- Communicating, reviewing, and coordinating directly with Program Panels to improve the information available to the panels for critical decision-making.
- Using project-provided descoping or rescoping data to assess mission costs and technical risks.
- Scheduling to preclude TRACE from becoming the pacing schedule item and allowing for the TRACE results to further inform the steering committee deliberations and decision-making.
- Using varied budget and funding projections, nominal and optimistic, to frame project execution opportunities and challenges.
- Factoring costs of current and near-available launch vehicles into mission costs.
- Using experts, as needed, to augment the Aerospace Corporation independent assessment process, after gaps were identified.
- Identifying liaisons to serve as go-betweens the panels, the committee, and the Aerospace Corporation's TRACE team.
- Early identification of the role of TRACE in panel decisions and deliberations.
- Presentations of final TRACE results directly to the Program Panels.

The overall process is captured in Figure O.2. Science priorities were provided to the Program Panels as input to the evaluation process. An early assessment of project science, scope, technical risks, costs, possible descopes, and associated risks were used as input. The Program Panels used requests for

---

[1] National Academies of Sciences, Engineering, and Medicine (NASEM), 2015, *The Space Science Decadal Surveys: Lessons Learned and Best Practices*, The National Academies Press, Washington, D.C.

[2] D. Emmons, 2016, "Aerospace CATE Overview: Astro2020 Large Mission Concepts," teleconference presentation to the LUVOIR Science and Technology Definition Team on September 20, 2016, https://asd.gsfc.nasa.gov/luvoir/events/events/telecons/2016-09-20/emmons.pptx.

[3] National Research Council, 2013, *Lessons Learned in Decadal Planning in Space Science: A Summary of a Workshop*, The National Academies Press, Washington, D.C.

[4] NASEM, 2015, *The Space Science Decadal Surveys: Lessons Learned and Best Practices*, The National Academies Press, Washington, D.C.

information (RFI) to gather data on projects using a two-tiered process. Projects of greater maturity were sent RFIs reflecting more detailed data (RFI1). Projects of lesser maturity were sent an alternate, lesser detailed RFI (RFI2).

Outcomes were presented to the Program Panels using standardized formats. Results included technical risk ratings, power and mass margins, and other risk data. Cost histograms were presented to compare costs developed using a range of independent cost estimating methods, including classical estimating methods, risk estimating programs (e.g., SEER, MICM, NICM, MOCET, FRISK, etc.[5]), and the contractor's proprietary analytics suite. The key analogies and historical projects used in the analyses for the various design elements, subsystems, instruments, construction elements, etc., were also reviewed by the Program Panels to assess appropriate application and allow for programs of varied maturities to be brought to a consistent baseline. TRACE products provided to the Program Panels during the process flow are shown in Figure O.3.

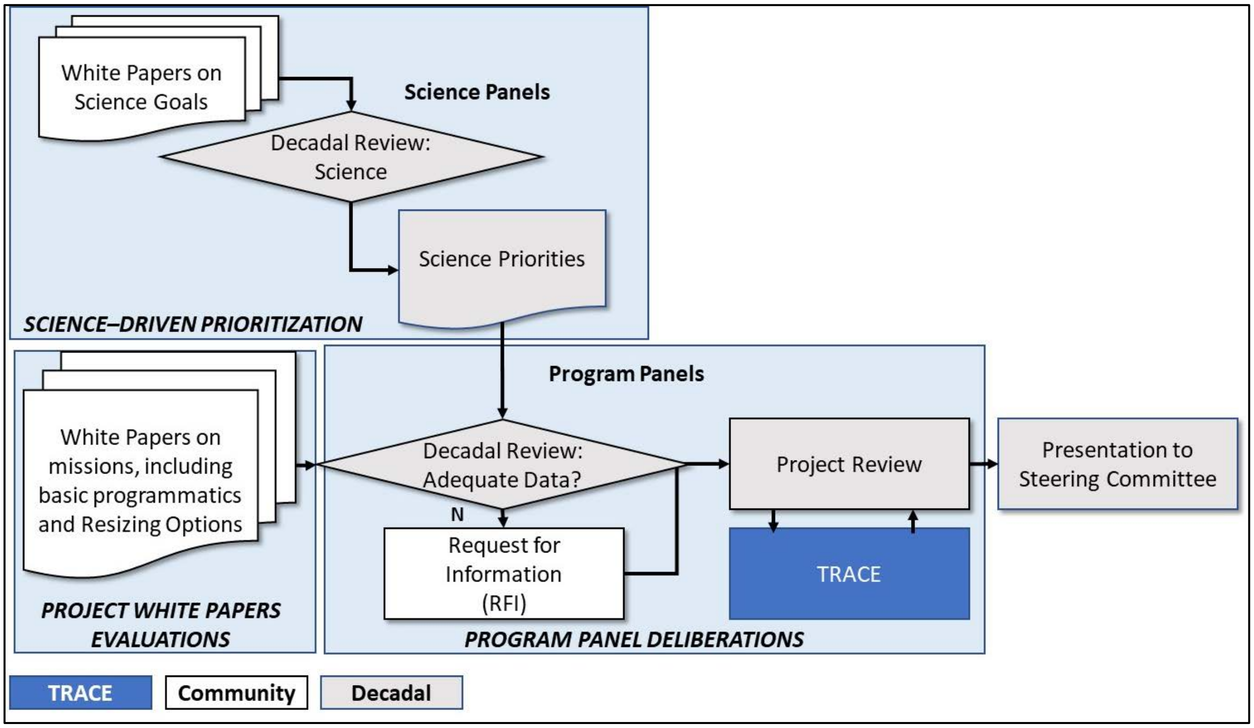


FIGURE O.2 Technical Risk and Cost Evaluation (TRACE) process interactions.

[5] A subset of classic estimating tools used by TRACE:

- SEER (Systems Evaluation and Estimation of Resources), a knowledge-based estimation cost evaluation and trade-off tool suite by Galorath, Inc.
- MICM (Multi-variable Instrument Cost Model), by NASA Goddard Space Flight Center.
- NICM (NASA Instrument Cost Model), a probabilistic cost estimation tool developed by NASA containing a system model, subsystem model, and a database search engine.
- MOCET (Mission Operations Cost Estimation Tool), a Phase E estimation tool jointly developed by The Aerospace Corporation and the NASA Science Office for Mission Assessments.
- PCEC (Project Cost Estimating Capability), a parametric cost model developed and maintained by NASA.
- FRISK (Formal RISK method), an analytical, rather than Monte Carlo–based, risk analysis method. Typical implementation includes establishing total cost distribution for a design based on WBS (Work Breakdown Structure).

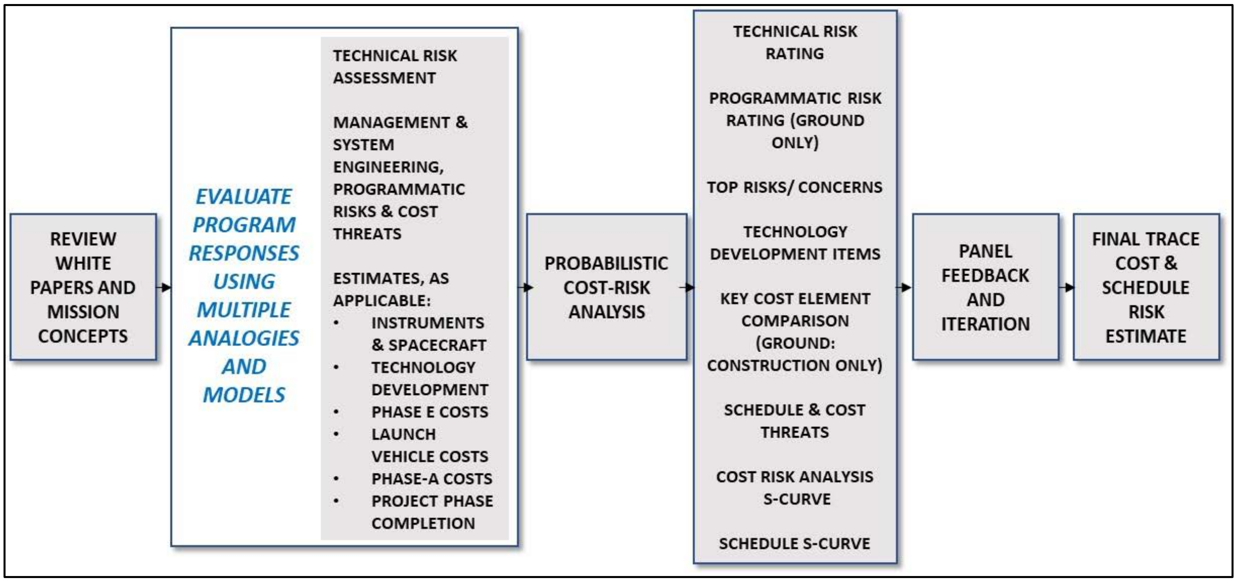


FIGURE O.3 TRACE process and products.

Space concepts benefitted from the extensive past program database available to Aerospace Corporation allowing for numerical analytical methods to be effectively and consistently applied across a range of concepts. However, the unique aspects of ground observatories continued to provide challenges to the TRACE. Aerospace made significant efforts to improve the TRACE process for ground programs including adding additional historical analogy data and updating their analyses to include current applicable facilities. Ground TRACE evaluations limitations included the exclusion of operations costs, focusing only on design and construction, and the exclusion of funding threats. All gaps were addressed by the Program Panels, engaging additional unpaid consultants where necessary to augment the database and evaluate program specifics.

The TRACE processes forecasted final project costs, timelines, and critical paths, allowing “unknown unknowns,” or “threats,” to be captured, typically producing more conservative values, e.g., manufacturing development and scale-up, increased schedule to reflect demonstrated development timelines, added costs to capture mass contingencies, increased costs, and time to reflect required launch vehicle capacity and availability, increased costs and schedule to capture historic execution and learning curves associated with specific technologies, etc. Once overall costs and schedules were estimated, probabilistic assessments of cost probability and schedule probability were calculated and provided to the Program Panels using classic S-curves comparing project estimates to 70th percentile probabilities.

The role of the TRACE data as input to panel decision making is presented in the individual Program Panel reports.

The projected cost growth range for large space and ground mission concepts in real year dollars is shown in Figure O.4. It should be emphasized that the “Increased projected cost” in the figure reflects the differences between the TRACE total mission cost estimates at 70 percent probability and the project total mission cost estimates. The differences in projected growth assigned to ground versus space concepts were generally attributable to technology maturity gaps and the demonstrated cost and scope growth challenges faced by early phase projects.

The impact of early phase programmatic risk on anticipated program cost growth is shown in Figure O.5. (Increased technical risk is associated with higher projected cost growth.) An examination of the aggregate TRACE results, as a percentage of the total project value, were compared to the open technical risks. Every project that received a TRACE evaluation was assessed for technical risk by evaluating both mitigation plans and historical risks. Figures of merit (FOM) were used to normalize the data and highlight relationships. FOM formulae are provided in Box O.1. In addition to the general trend

of higher technical risks driving forecasted cost growth, a distinct difference was apparent between the inherent technical risks of the space mission concepts and the ground concepts.

FIGURE O.4 TRACE projected cost growth (70 percent probability level).

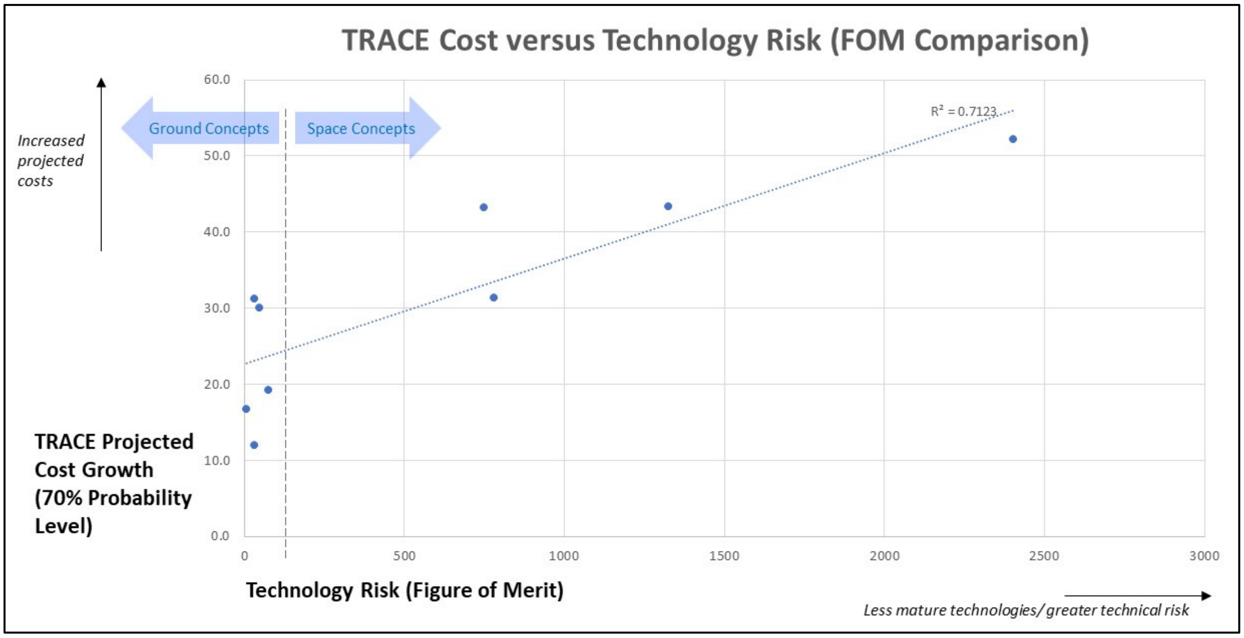


FIGURE O.5 Correlation of TRACE “look forward” to technical risk.

BOX O.1 Figures of Merit Calculations

To assess the extent to which the calculated program execution threats (TRACE) were correlated to open technical risks across all programs, figures of merit (FOM) were developed. TRACE values were normalized as a percentage of project values:

$TR_{project} = ((TR_{cost} \times 100)/ P_{cost}) - 100$, where

- $TR_{project}$ = TRACE percent increase over total project submitted cost
- TR_{cost} = TRACE estimate for complete project in real year dollars
- P_{cost} = Project estimate for complete project in real year dollars

As an example, if a project cost totaled $1,000 million in real year dollars and the TRACE 70 percent probability value was $1,100 million in real year dollars, $TR_{project} = 10$. Because the TRACE process captures and monetizes risks, all values of $TR_{project}$ were positive.

Technical risks were normalized using a FOM capturing the number of large open technical risk categories, the technology readiness level (TRL) of the open technology, and the time (in years) needed to mature the technology as submitted by the program:

$$\begin{aligned} Tech_{FOM} = {} & (R_{high} \times (TRL3 \times TRL3_{time} + TRL4 \times TRL4_{time} + TRL5 \times TRL5_{time} + TRL6) + \\ & (R_{medium\text{-}high} \times (TRL3 \times TRL3_{time} + TRL4 \times TRL4_{time} + TRL5 \times TRL5_{time} + TRL6) + \\ & (R_{medium} \times (TRL3 \times TRL3_{time} + TRL4 \times TRL4_{time} + TRL5 \times TRL5_{time} + TRL6) + \\ & (R_{medium\text{-}low} \times (TRL3 \times TRL3_{time} + TRL4 \times TRL4_{time} + TRL5 \times TRL5_{time} + TRL6) + \\ & (R_{low} \times (TRL3 \times TRL3_{time} + TRL4 \times TRL4_{time} + TRL5 \times TRL5_{time} + TRL6), \text{ where} \end{aligned}$$

- $_{TechFOM}$ = Technology Risk FOM
- R_{high}, $R_{medium\text{-}high}$, R_{medium}, $R_{medium\text{-}low}$, and R_{low} = Technology risk rating assigned by Aerospace = Number of major program risks in this risk category × Risk multiple, where
  - — High risk multiple = 5
  - — Medium-high risk multiple = 4
  - — Medium risk multiple = 3
  - — Medium-low risk multiple = 2
  - — Low risk multiple = 1
- TRL3, TRL4, TRL5, and TRL6 = number of technology categories identified by the project at these TRL levels × TRL risk multiplier, where
  - — TRL 3 multiplier = 9
  - — TRL 4 multiplier = 3
  - — TRL 5 multiplier = 2
  - — TRL 6 multiplier = 1
- $TRL3_{time}$, $TRL4_{time}$, and $TRL5_{time}$ = total length of time, in years as defined by the projects for a technology to mature to the next level.

As an example, if a project defined two major development areas, one currently at TRL 6 and rated a medium risk by Aerospace Corporation, and the other at TRL 3 and rated a medium-low risk by Aerospace Corporation, but requiring 2 years each to reach TRL 4, TRL 5, and TRL6, the project $Tech_{FOM} = 61$.

P

Acronyms

4MOST	4-metre Multi-Object Spectroscopic Telescope
A&G	acquisition and guiding
AAAC	Astronomy and Astrophysics Advisory Committee
AAAS	American Association for the Advancement of Science
AAG	Astronomy and Astrophysics Research Grants
AAS	American Astronomical Society
ACT	Atacama Cosmology Telescope
ACTPOL	Atacama Cosmology Telescope Polarization Survey
ADA	Americans with Disabilities Act
ADAP	Astrophysics Data Analysis Program
ADAS	Astronomical Data Archiving System
AEON	Astronomical Event Observatory Network
AGB	asymptotic giant branch
AGEP	Alliances for Graduate Education and the Professoriate
AGN	active galactic nuclei
AIP	American Institute of Physics
AIPS	Astronomical Image Processing System
ALMA	Atacama Large Millimeter/Submillimeter Array
AM CVn	Canum Venaticorum
AMANDA	Antarctic Muon and Neutrino Detector Array
AMEGO	All-sky Medium Energy Gamma-ray Observatory
AMO	atomic, molecular, and optical
ANITA	Antarctic Impulsive Transient Antenna
AO	adaptive optics
AO	announcement of opportunity
AON	Arctic Observing Network
APC	activity, project, or state of the profession white paper
APL	Applied Physics Laboratory
APOGEE	Apache Point Observatory Galactic Evolution Experiment
APRA	Astrophysics Research and Analysis Program
APS	American Physical Society
APT	Advanced Particle-astrophysics Telescope
ARA	Askaryan Radio Array
ARC	Astrophysics Research Consortium Telescope
ARGOS	Advanced Rayleigh Guided Ground Layer Adaptive Optics System
ARIANNA	Antarctic Ross Ice-Shelf Antenna Neutrino Array
ARIEL	Atmospheric Remote Sensing Infrared Exoplanet Large-Survey
ASAS-SN	All-Sky Automated Survey for Supernovae
ASCR	Advanced Scientific Computing Research
ASKAP	Australian Square Kilometre Array Pathfinder

ASM	adaptive secondary mirror
AST	NSF Division of Astronomical Sciences
Athena	Advanced Telescope for High Energy Astrophysics
ATI	Advanced Technologies and Instrumentation
ATP	Astrophysics Theory Program
AU	astronomical unit
AUI	Associated Universities, Inc.
AURA	Association of Universities for Research in Astronomy
AXAF	Advanced X-Ray Astrophysics Facility
AXIS	Advanced X-Ray Imaging Satellite
BAO	baryon acoustic oscillations
BBN	big bang nucleosynthesis
BH	black hole
BICEP	Background Imaging of Cosmic Extragalactic Polarization
BOINC	Berkeley Open Infrastructure for Network Computing
CALIFA	Calar Alto Legacy Integral Field Area
CASTOR	Cosmological Advanced Survey Telescope for Optical and UV Research
CATE	Cost and Technical Evaluation
CCAT	Cerro Chajnantor Atacama Telescope
CCD	charge-coupled device
CDM	cold dark matter
CE	Cosmic Explorer
CERN	European Council for Nuclear Research
CETUS	Cosmic Evolution Through UV Surveys
CGI	computer generated imagery
CGM	circumgalactic medium
CHARA	Center for High Resolution Astronomy
CHEOPS	CHaracterizing ExOPlanet Satellite
CHIME	Canadian Hydrogen Intensity Mapping Experiment
CHORD	Canadian Hydrogen Observatory and Radio-transient Detector
CLASS	Cosmology Large Angular Scale Surveyor
CMB	Cosmic Microwave Background
CMB-S4	Cosmic Microwave Background Stage 4
CME	Coronal Mass Ejections
CMOS	complementary metal oxide semiconductor imaging sensors
CMS	concept maturation studies
CNES	Centre National d'Etudes Spatiales
CNM	cold neutral medium
CORF	Committee on Radio Frequencies
CoRoT	Convection Rotation and Planetary Transits
COS	Cosmic Origins Spectrograph
COSI	Compton Spectrometer and Imager
COSMO	COronal Solar Magnetism Observatory
COVID-19	Coronavirus Disease
CPD	circumplanetary disk
CPU	central processing unit
CSBF	Columbia Scientific Balloon Facility
CSDC	Community Science and Data Center
CSMA	Committee on the Status of Minorities in Astronomy

CSO	Caltech Submillimeter Observatory
CSWA	Committee on the Status of Women in Astronomy
CTA	Cherenkov Telescope Array
CTE	coefficient of thermal expansion
CTIO	Cerro Tololo Inter-American Observatory
DA	discovery area
DECam	Dark Energy Camera
DESI	Dark Energy Spectroscopic Instrument
DFW	drop, fail, withdrawal
DKIST	Daniel K. Inouye Solar Telescope
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V. (German Aerospace Center)
DM	dark matter
DMR	Division of Materials Research
DMS	Division of Mathematical Sciences
DOE	Department of Energy
DRM	design reference mission
DSA-2000	Deep Synoptic Array 2000
DSOC	deep space optical communications
DST/HCIT	Decadal Survey Testbed/High Contrast Imaging Testbed
EBIO	education, broader impacts, and outreach
EDGES	Experiment to Detect the Global EoR Signature
E-ELT	European Extremely Large Telescope
EGS	Extended Groth Strip
EHT	Event Horizon Telescope
ELT	Extremely Large Telescope
EM	electromagnetic
EoR	epoch of reionization
EOVSA	Expanded Owens Valley Solar Array
EP3	Effective Practices for Physics Programs
EPRV	extreme precision radial velocity
eROSITA	Extended Roentgen Survey with an Imaging Telescope Array
ESA	European Space Agency
ESCAPE	Extreme-ultraviolet Stellar Characterization for Atmospheric Physics and Evolution
ESDIS	Earth Science Data and Information System
ESM	Electromagnetic Spectrum Management Group
ESO	European Southern Observatory
ESS	Exoplanet Science Strategy
ETL	Exoplanet Technology Laboratory
Euclid DF-Fornax	Euclid Deep Field Fornax
Euclid DF-S	Euclid Deep Field South
Euclid NEP	Euclid North Ecliptic Pole
EUV	extreme ultraviolet
EVLA	Expanded Very Large Array
Far-IR	far-infrared
FASR	Frequency Agile Solar Radiotelescope
FaST	Faculty and Student Teams
FAST	Five-hundred-meter Aperture Spherical Telescope

FBOT	fast and blue optical transients
FCC	Federal Communications Commission
FDSS	Faculty Development in Space Sciences
FIP	Far-IR Imager Polarimeter
FMR	SOFIA Five Year Flagship Mission Review
FOBOS	Fiber-Optic Broadband Optical Spectrograph
FOV	field of view
FRB	fast radio burst
FTE	full-time employee
FTS	fourier transform spectrometer
FUV	far ultraviolet
FY	fiscal year
GADGET	Galaxies with Dark Matter and Gas Interact
GALAH	Galactic Archeology with HERMES
GBO	Green Bank Observatory
GBT	Green Bank Telescope
GC	globular cluster
G-CLEF	GMT Consortium Large Earth Finder
GECAM	Gravitational Wave High-energy Electromagnetic Counterpart All-sky Monitor
GeMS	Gemini Multi-Conjugate Adaptive Optics System
GEP	Galaxy Evolution Probe
GeV	gigaelectronvolt
GLAO	ground layer adaptive optics
GMACs	GMT Multi-object Astronomy and Cosmology Spectrograph
GmagAO-X	Giant Magellan Telescope Extreme Adaptive Optics
GMT	Giant Magellan Telescope
GMTIFS	GMT Integral Field Spectrograph
GMTNIRS	GMT Near-IR Spectrograph
GNAO	Gemini North Adaptive Optics
GO	guest observer
GOMMTD	Great Observatory Mission Maturation and Technology Development
GONG	Global Oscillations Network Group
GOODS-N	Great Observatories Origins Deep Survey-North
GOODS-S	Great Observatories Origins Deep Survey-South
GPI	Gemini Planet Imager
GPS	global positioning system
GPU	graphical processing unit
GR	general relativity
GRB	gamma ray burst
GRE	Graduate Record Examination
GR-MHD	general-relativistic magnetohydrodynamics
GST	Goode Solar Telescope
GUSTO	Galactic/Extragalactic ULDB Spectroscopic Terahertz Observatory
GW	gravitational waves
GWOSC	Gravitational Wave Open Science Center
HabEx	Habitable Exoplanet Observatory
HAWC	High-Altitude Water Cherenkov Observatory
HBCU	historically black colleges and universities
HXDI	High Definition X-ray Imager

HE	high energy
HEASARC	High Energy Astrophysics Science Research Center
HEP	High Energy Physics
HERA	Hydrogen Epoch of Reionization Array
HET	Hobby-Eberly Telescope
HIRAX	Hydrogen Intensity and Real-time Analysis Experiment
HOSTS	Hunt for Observable Signatures of Terrestrial Systems
HPC	high-performance computing
HPD	half power diameter
HR	Hertzprung-Russell
HST	Hubble Space Telescope
HZ	habitable zone
IACT	Imaging Atmospheric Cherenkov Telescope
IAU	International Astronomical Union
IceCube-Gen2	IceCube Generation 2
ICM	intracluster medium
IFU	integral field unit
IGM	intergalactic medium
IGW	inflationary gravitational waves
IMBH	intermediate mass black holes
IMF	initial mass function
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
IPAC	Infrared Processing and Analysis Center
IPTA	International Pulsar Timing Array
IR	infrared
IRAF	Image Reduction and Analysis Facility
IRIS	Infrared Imaging Spectrometer
IR/O/UV	infrared/optical/ultraviolet
ISM	interstellar medium
ISO	Infrared Space Observatory
ITU	International Telecommunications Union
IXPE	Imaging X-ray Polarimetry Explorer
JAXA	Japan Aerospace Exploration Agency
JHU	John Hopkins University
JPL	Jet Propulsion Laboratory
JVLA	Karl Jansky Very Large Array
JWST	James Webb Space Telescope
KAGRA	Kamioka Gravitational-Wave Detector
KAPA	Keck All-Sky Precision Adaptive Optics
KBO	Kuiper Belt object
KLIP	Karhunen-Lo`eve image processing
KPIC	Keck Planet Imager and Characterizer
KPNO	Kitt Peak National Observatory
LAT	Large Aperture Telescope
LBA	Long Baseline Array
LBT	Large Binocular Telescope
LBTI	Large Binocular Telescope Interferometer

LBTO	Large Binocular Telescope Observatory
LCIT	Large Mission Concept Independent Assessment Team
LCOGT	Las Cumbres Observatory Global Telescope
LDB	long duration balloon
LEAP	Large Area Burst Polarimeter
LEDA	Large-Aperture Experiment to Detect the Dark Ages
LEO	low Earth orbit
LGBTQIA+	Lesbian, Gay, Bisexual, Trans, Queer, Intersex, Asexual, Plus
LHAASO	Large High Altitude Air Shower Observatory
LHC	Large Hadron Collider
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
LMA	Lynx Mirror Assembly
LMC	Large Magellanic Cloud
LMT	Large Millimeter Telescope
LSST	Legacy Survey of Space and Time
LTAO	laser tomography adaptive optics
LTK	local and traditional knowledge
LUVOIR	Large UV/Optical/IR Surveyor
LXM	Lynx X-ray Microcalorimeter
LyC	Lyman continuum
MAGIC	Major Atmospheric Gamma Imaging Cherenkov Telescope
MaNGA	Mapping nearby Galaxies at APO
MARC	Maximizing Access to Research Careers
MCAO	multi-conjugate adaptive optics
MEMS	micro-electro-mechanical-system
MESA	Modules for Experiments in Stellar Astrophysics
MEV	Maximum Expected Value
MHD	magnetohydrodynamics
MIDEX	Medium-class Explorers
MINERVA-Australis	Miniature Exoplanet Radial Velocity Array-Australis
MIR	mid-infrared
MIRC	Michigan Infrared Beam Combiner
MISC-T	Mid-infrared Spectrometer Camera Transit Spectrometer
MKID	microwave kinetic induction detectors
MKO	Mauna Kea Observatory
MKSR	Mauna Kea Science Reserve
MLSO	Mauna Loa Solar Observatory
MMA	multi-messenger sstrophysics
MMT	Multiple Mirror Telescope
MO and MOO	missions of opportunity
MOAO	multi-object adaptive optics
MODHIS	Multi-Object Diffraction-limited High-Resolution Infrared Spectrograph
MPE	Max Planck Institute for Extraterrestrial Physics
MPS	Directorate for Mathematical and Physical Sciences
MREFC	Major Research Equipment and Facilities Construction
MRI	Major Research Instrumentation
MROI	Magdalena Ridge Optical Interferometer
MRSEC	Materials Research Science and Engineering Centers

MSE	Maunakea Spectroscopic Explorer
MSFC	Marshall Space Flight Center
MSI	minority serving institution
MSIP	Mid-Scale Innovations Program
MSO	Mid-Scale Opportunities
MSRI	Mid-Scale Research Infrastructure
MUCERPI	Minority University and College Education and Research Partnership Initiative
MW	Milky Way
MWA	Murchison Widefield Array
NANOGrav	North American Nanohertz Observatory for Gravitational Waves
NASA	National Aeronautics and Space Administration
NEID	NN-EXPLORE Exoplanet Investigations with Doppler Spectroscopy
NEOWISE	Near-Earth Object Wide-field Infrared Survey Explorer
NICER	Neutron Star Interior Composition Explorer
NFIRAOS	narrow field infrared adaptive optics system
ngVLA	Next-Generation Very Large Array
NICER	Neutron Star Interior Composition Explorer
NIH	National Institutes of Health
NIR	near-infrared
NIST	National Institute of Standards and Technology
NLSO	NASA LISA Studies Office
NN-EXPLORE	NASA-NSF Exoplanet Observational Research program
NOAO	National Optical Astronomy Observatory
NOIRLab	National Optical-Infrared Research Laboratory
NPOI	Navy Precision Optical Interferometer
NRAO	National Radio Astronomy Observatory
NS	neutron star
NS EOS	neutron star equation of state
NSB	National Science Board
NSBP	National Society of Black Physicists
NSF	National Science Foundation
NSHP	National Society of Hispanic Physicists
NSO	National Solar Observatory
NSPIRES	NASA Solicitation and Proposal Integrated Review and Evaluation System
NuSTAR	Nuclear Spectroscopic Telescope Array
NVO	National Virtual Observatory
NWNH	New Worlds, New Horizons report
O&M	operations and management
OIR	optical and infrared
ORM	Roque de los Muchachos Observatory
OSS	open source software
OSS	Origins Survey Spectrometer
P5	Particle Physics Project Prioritization Panel
PAARE	Partnerships in Astronomy and Astrophysics Research
PAG	program analysis groups
PAH	polycyclic aromatic hydrocarbons
PAMS	Portfolio Analysis and Management System
Pan-STARRS	Panoramic Survey Telescope and Rapid Response System

pc	parallax of one arc-second (parsec)
PDR	preliminary design review
PER	Physics Education Research
PFS	Prime Focus Spectrograph
PGRE	Physics Graduate Record Examination
PHAT	Panchromatic Hubble Andromeda Treasury Project
PI	principal investigator
PIC	particle-in-cell
PICASO	Planetary Intensity Code for Atmospheric Scattering
PICTURE	Planetary Imaging Concept Testbed Using a Recoverable Experiment
PISN	pair-instability supernovae
PLATO	Planetary Transits and Oscillations
POEMMA	Probe of Extreme Multi-Messenger Astrophysics
POLARBEAR	Polarization of Background Radiation
PTA	pulsar timing arrays
PTF	Palomar Transient Factory
PUI	primarily undergraduate institution
PUMA	Packed Ultra-wideband Mapping Array
QCD	quantum chromodynamics
QIS	quantum information science
QSO	quasi-stellar object
R	resolution
RBIS	research-based instructional strategies
R&D	Research and development
REU	Research Experiences for Undergraduates
RFI	radio grequency interference
RFI	request for information
RICE	Radio Ice Cherenkov Experiment
RMS	radio, millimeter, and submillimeter
RMS	root mean square
RNO-G	Radio Neutrino Observatory in Greenland
Roman ST HLS	Nancy Grace Roman Space Telescope High Latitude Survey
ROSAT	Roentgen Satellite
ROSES	Research Opportunities in Space and Earth Sciences
RV	radial velocity
RXTE	Rossi X-Ray Timing Explorer
SACNAS	Society for Advancement of Chicanos/Hispanics and Native Americans in Science
SAFARI	Spica FAR Infrared Instrument
SALT	South African Large Telescope
SAO	Smithsonian Astrophysical Observatory
SAT	Small Aperture Telescope
SAT	Strategic Astrophysics Technology
SATCON	Working Group on Satellite Constellations
SCExAO	Subaru Coronagraphic Extreme Adaptive Optics
SCT	Schwartzchild-Couder IACT Telescope
SDO	Solar Dynamics Observatory
SDSS	Sloan Digital Sky Survey
SEA Change	STEMM Equity Achievement Change program

SED	spectral energy distribution
SETI	Search for Extraterrestrial Intelligence
SFR	star formation rate
SGMA	Committee for Sexual-Orientation and Gender Minorities in Astronomy
SIM	Athena Science Instruments Module
SIRTF	Space Infrared Telescope Facility
SKA	Square Kilometre Array
SLSNe	super-luminous supernovae
SMA	Submillimeter Array
SMARTS	Small and Moderate Aperture Research Telescope System
SMBH	supermassive black hole
SMC	Small Magellanic Cloud
SMD	Science Mission Directorate
SMEX	Small Explorers
SNe	supernovae
SO	Simons Observatory
SOAR	Southern Astrophysical Research Telescope
SOFIA	Stratospheric Observatory for Infrared Astronomy
SOLIS	Synoptic Optical Long-term Investigations of the Sun
SOMER	SOFIA Operations and Maintenance Efficiency Review
SPB	super pressure balloon
SpecTel	Spectroscopic Survey Telescope
SPHERE	Spectro-Polarimetric High-contrast Exoplanet Research Instrument
SPHEREx	Spectro-Photometer for the History of the Universe, Epoch of Reionization and Ices Explorer
SPICA	Space Infrared Telescope for Cosmology and Astrophysics
SPO	South Pole Observatory
SPT	South Pole Telescope
SQL	structured query language
SQUID	superconducting quantum interference devices
SRON	Netherlands Institute for Space Research
S-STEM	Scholarships in Science, Technology, Engineering, and Mathematics
STDT	science and technology definition team
STEM	Science, Technology, Engineering, and Mathematics
STIS	Space Telescope Imaging Spectrograph
STMD	Space Technology Mission Directorate
STScI	Space Telescope Science Institute
SVOM	Space Variable Objects Monitor
SWGO	Southern Wide-Field Gamma-Ray Observatory
SZ	Sunyaev-Zel'dovich (effect)
TAP	Transient Astrophysics Probe
TCAN	theory and computation network
TCU	tribal colleges and universities
TDA	time domain astronomy
TDE	tidal disruption event
TEAM-UP	National Task Force to Elevate African American Representation in Undergraduate Physics & Astronomy
TES	transition edge sensor
TESS	Transiting Exoplanet Survey Satellite
TIO	Thirty Meter Telescope International Observatory

TK	Indigenous traditional knowledge
TMT	Thirty Meter Telescope
TOMS	target and observation managers
TRACE	technical, risk, and cost evaluation
TRL	technology readiness level
TSIP	Telescope System Instrumentation Program
TTV	transit timing variations
UHE	ultra-high energy
UHECR	ultra-high-energy cosmic rays
ULDB	ultra-long duration balloon
ULTRA	Ultra-stable Large Telescope Research and Analysis
ULX	ultraluminous X-ray
URO	University Radio Observatories
USRA	Universities Space Research Association
UV	ultraviolet
UVOT	UV-optical telescope
VFP	Visiting Faculty Program
VHE	very high energy
VLA	Very Large Array
VLASS	Very Large Array Sky Survey
VLBA	Very Long Baseline Array
VLBI	very long baseline interferometry
VLT	Very Large Telescope
VLTI	Very Large Telescope Interferometer
WC	women’s college
WD	white dwarf
WEAVE	William Herschel Telescope (WHT) Enhanced Area Velocity Explorer
WFI	Wide Field Imager
WFIRST	Wide-Field Infrared Survey Telescope
WFS	wave-front sensors
WGAD	Working Group on Accessibility and Disability
WIMPS	weakly interacting massive particles
WISE	Wide-Field Infrared Survey Explorer
WIYN	Wisconsin-Indiana-Yale-NOAO Observatory
WMAP	Wilkinson Microwave Anisotropy Probe
WNM	warm neutral medium
WRC	World Radio Communication Conference
XGS	Lynx X-Ray Grating Spectrometer
X-IFU	X-Ray Integral Field Unit
XMM-Newton	X-Ray Multi Mirror Mission
XRISM	X-Ray Imaging and Spectroscopy Mission
XRP	Exoplanet Research Program
ZDI	Zeeman Doppler Imaging
ZTF	Zwicky Transient Facility

Q

Committee and Panel Biographical Information

STEERING COMMITTEE

FIONA A. HARRISON, *Co-Chair,* is the Benjamin M. Rosen Professor of Physics and the Kent and Joyce Kresa Leadership Chair of the Division of Physics and Mathematics at the California Institute of Technology, Pasadena. Dr. Harrison's primary research interests are in experimental and observational high-energy astrophysics. She is the principal investigator of NASA's Nuclear Spectroscopic Telescope Array (NuSTAR), for which she received the NASA Outstanding Public Leadership Medal in 2013. In 2015, Dr. Harrison was awarded the Bruno Rossi Prize of the High Energy Astrophysics Division of the American Astronomical Society, and in 2016 she won the Harrie Massey Award from the Committee on Space Research. She was elected to the American Academy of Arts and Sciences and the National Academy of Sciences in 2014. Dr. Harrison is past chair of the Division of Astrophysics of the American Physical Society, and chair-elect of the High Energy Astrophysics Division of the American Astronomical Society. Dr. Harrison served as chair of the National Academies of Sciences, Engineering, and Medicine's Space Studies Board, is a member of the James Webb Space Telescope Independent Review Board, and chaired the National Academies' Committee on an Assessment of the Astrophysics Focused Telescope Assets (AFTA) Mission Concepts. She was a member of the Committee on Decadal Survey on Astronomy and Astrophysics 2010.

ROBERT C. KENNICUTT JR., *Co-Chair,* is a professor at the Steward Observatory at the University of Arizona and in the Department of Physics and Astronomy at Texas A&M University. His research interests are primarily in observational extragalactic astronomy and cosmology. Dr. Kennicutt has over 40 years of experience in various capacities, including serving as Plumian Professor of Astronomy and Experimental Philosophy and as director of the Institute of Astronomy, and as head of the School of Physical Sciences at the University of Cambridge; as editor-in-chief of the *Astrophysical Journal*; and as professor/astronomer and deputy head of the Department of Astronomy at the University of Arizona. Dr. Kennicutt has won numerous awards, including the Gruber Cosmology Prize and the Dannie Heinman Prize in Astrophysics at the American Institute of Physics. He received his Ph.D. in astronomy from the University of Washington. He was elected to the National Academy of Sciences in 2006 and to the American Academy of Arts and Sciences in 2001 and was elected a fellow of the Royal Society of London (FRS) in 2011. Dr. Kennicutt has served on various committees at the National Academies, including the Committee on Decadal Survey on Astronomy and Astrophysics 2010, the Committee on Astronomy and Astrophysics and the Task Group on Space Astronomy and Astrophysics.

JULIANNE DALCANTON is a professor and chair of astronomy at the University of Washington. Her research interests include the origin and evolution of galaxies and their use as probes of fundamental physics. Dr. Dalcanton is also the principal investigator of a large Hubble Space Telescope Multi-Cycle Treasury, has served as the vice chair of the Space Telescope Science Institute Council, as a member of the Collaboration Council of the Sloan Digital Sky Survey (SDSS), as the chair of the SDSS Galaxy Working Group, and as a member of the Association of Universities for Research in Astronomy (AURA) nominating committee. Prior to joining her current institution, she was a postdoctoral fellow at the

Observatories of the Carnegie Institution of Washington. Dr. Dalcanton is the recipient of numerous awards, including the Alfred P. Sloan Research Fellowship, a National Science Foundation CAREER Award for beginning faculty, a NASA Hubble Postdoctoral Fellowship, a Wyckoff Faculty Fellowship through the College of Arts and Sciences at the University of Washington, the Mohler Prize from University of Michigan, and the Beatrice Tinsley Prize from the American Astronomical Society. She received her Ph.D. in astrophysical sciences from Princeton University. Dr. Dalcanton served on the National Academies' Astro2010 Decadal Survey Panel on the Galactic Neighborhood.

TIM DE ZEEUW is a professor of astronomy at Leiden University. Dr. De Zeeuw also holds a senior visiting position at the Max Planck Institute for Extraterrestrial Physics. His research interests include the formation, structure, and dynamics of galaxies, including the Milky Way. Dr. De Zeeuw has previously served as director general of the European Southern Observatory, and as director of the Leiden Observatory and of the Netherlands Research School for Astronomy. He has led the development of the European science vision for astronomy. Prior to joining Leiden University, Dr. De Zeeuw was a senior research fellow at the California Institute of Technology, a long-term member of the Institute for Advanced Study, and a teaching and research assistant at Leiden University. He has received numerous awards, including the Royal Astronomical Society Group Award for the Spectrographic Areal Unit for Research on Optical Nebulae (SAURON) Team and the Brouwer Award of the Dynamical Division of the American Astronomical Society. Dr. De Zeeuw is a member of the Royal Netherlands Academy of Arts and Sciences, the European Astronomical Society, the AAS, and the International Astronomical Union. He received his Ph.D. in astronomy from the University of Leiden.

ANDREW S. DRIESMAN is a member of the principal professional staff in the Space Sector of the Johns Hopkins University Applied Physics Laboratory (JHU APL). His background and experience are in program management, organizational management, systems engineering, integration, and architecting of complex spacecraft for both scientific and military use. Dr. Driesman is currently the program manager for NASA's Parker Solar Probe mission. Prior to starting his current role, he served in various positions including technical director of the Joint Polar Satellite System at the National Oceanic and Atmospheric Administration (NOAA), and supervisor for the Space Systems Applications Group at JHU APL. Previously, Dr. Driesman served as the lead engineer for developing both Solar Terrestrial Relations Observatory (STEREO) spacecrafts, from conceptual design through on-orbit operations. Additional experience includes system engineering for military satellite systems, board-level analog designs for space shuttle payloads, board and subsystem-level design for balloon and sounding rocket payloads, and systems-level design for missile payloads and satellites. Dr. Driesman is the recipient of numerous awards, including the NASA Individual Achievement Award in 2008 and the NASA Exceptional Public Service Medal for Outstanding System Engineering Leadership Award. He received an M.S. in technical management from Johns Hopkins University.

JONATHAN J. FORTNEY is the director of the Other World Laboratory at the University of California, Santa Cruz (UC Santa Cruz). He is also a professor of astronomy and astrophysics. Prior to joining UC Santa Cruz, Dr. Fortney was a Spitzer Fellow with NASA Ames Research Center and a principal investigator at the SETI Institute. He also held a postdoctoral fellowship with the National Research Council at NASA Ames Research Center. Dr. Fortney's research interests include the interiors and atmospheres of planets in and out of the solar system, atmospheres and spectra of rocky and gas giant exoplanets, super Earth and giant planet thermal evolution, planetary interiors, exoplanet characterization through transit photometry and direct imaging, and the formation of giant planets. He has received numerous fellowships and awards, including the Urey Prize in the Division of Planetary Sciences with the American Astronomical Society, the 2010 Alfred P. Sloan Fellowship, the NASA Early Career Fellowship in Planetary Sciences, and as a National Academy of Sciences Kavli Fellow. Dr. Fortney received his Ph.D. in planetary science from the University of Arizona.

GABRIELA GONZÁLEZ is a professor of physics and astronomy at Louisiana State University (LSU). Dr. González is also the former spokesperson of the Laser Interferometer Gravitational-Wave Observatory (LIGO) Collaboration in the department of physics and astronomy. She is a leader of the LIGO collaboration to detect gravitational waves that successfully observed a signal on September 15, 2015, generated by the collision of a binary system of black holes. Prior to joining LSU, Dr. González was an assistant professor at Pennsylvania State University. She has received numerous honors and awards, including, most recently, the Southeastern Universities Research Association (SURA) Distinguished Scientist Award and the Dickinson College John Glover Award Medal. Dr. González received her Ph.D. in physics from Syracuse University. She is a member of the National Academy of Sciences and has served on the National Academies' Board on Higher Education and the Workforce and the Astro2010 Decadal Survey Panel on Particle Astrophysics and Gravitation.

JORDAN A. GOODMAN is a Distinguished University Professor of Physics at the University of Maryland. Dr. Goodman's research interests include particle astrophysics, which includes the study of cosmic radiation to better understand he properties in space that produce those particles, blending the elements of both high-energy physics and astrophysics. Dr. Goodman has served in various capacities at the University of Maryland, including former chair of the Physics Department. He is the principal investigator and has been the U.S. Spokesperson of the High-Altitude Water Cherenkov (HAWC) Gamma Ray Observatory. Dr. Goodman is the recipient of numerous awards, including the 2017 Yodh Prize for Astroparticle Physics Commission of IUPAP, the 2016 Breakthrough Prize in fundamental physics, and the University of Maryland President's Medal in 2009. He received his Ph.D. in physics from the University of Maryland.

MARC P. KAMIONKOWSKI is the William R. Kenan Jr. Professor of Physics and Astronomy at Johns Hopkins University. Dr. Kamionkowski is a theoretical physicist who specializes in cosmology, with contributions in dark matter, dark energy, the cosmic microwave background, the early universe, physical cosmology, along with other areas of astrophysics. Dr. Kamionkowski is also the chief editor for *Astrophysics* and a cosmology editor for *Physics Reports*. He is a fellow of the American Physical Society, the American Association for the Advancement of Science, the International Society for General Relativity and Gravitation, and the American Academy of Arts and Sciences and a member of the National Academy of Sciences. Dr. Kamionkowski has received numerous awards and honors, including the Helen B. Warner Prize, the E. O. Lawrence Award for Physics, a Simons Investigator Award, and the Dannie Heineman Prize for Astrophysics. He earned a Ph.D. in physics from the University of Chicago. Dr. Kamionkowski previously served on the National Academies' Astro2010 Panel on Cosmology and Fundamental Physics, the Panel on Theory and Computation in Astronomy and Astrophysics, and the Fifteenth Annual Symposium on Frontiers of Science.

BRUCE A. MACINTOSH is a professor of physics at Stanford University. His research focuses on the detection of extrasolar planets through direct imaging, and on development of adaptive optics and astronomical instrumentation for ground and space-based telescopes. Dr. Macintosh is a co-discoverer of four planets orbiting the star HR 8799 and is the principal investigator of the Gemini Planet Imager, an advance adaptive optics planet-finder for the Gemini South Telescope. Together with the HR8799 team, he received the 2009 Newcomb Cleveland Prize from the American Association for the Advancement of Science. He received his Ph.D. in astronomy at University of California, Los Angeles. Dr. Macintosh has served on the National Academies' Astro2010 Panel on Optical and Infrared Astronomy from the Ground, the Committee on Astronomy and Astrophysics, the Committee on Exoplanet Science Strategy, and the Committee on the Review of Progress Toward the Decadal Survey Vision in New Worlds, New Horizons in Astronomy and Astrophysics.

JACOBUS M. OSCHMANN is the 2019 president of the International Society for Optics and Photonics (SPIE). He retired from Ball Aerospace, where he had served as the vice president and general manager

of Civil Space. Mr. Oschmann is known for his significant contributions to the field of optical sciences, in optical design and technology development, along with his contributions and management on space and earth science instrumentation. He previously worked for the Association of Universities for Research in Astronomy (AURA), serving as project manager and chief engineer for the Gemini Observatory during its construction and early operations and then as the project manager for the conceptual design of the Daniel K. Inouye Solar Telescope. He has held various positions within SPIE, including on the board of directors, as past chair of SPIE Conference on Optical, Infrared and Millimeter Space Telescopes and SPIE Conference on Ground-Based and Airborne Telescopes. Mr. Oschmann currently serves on the NASA Advisory Subcommittee for Technology, Innovation, and Engineering and is chairing the first Atacama Large Millimeter Array (ALMA) Visiting Committee. He has also served on numerous review committees for NASA, the National Science Foundation, and AURA and the European Southern Observatory, including oversight and/or reviews for the JWST, the Large Synoptic Survey Telescope (LSST), National Solar Observatory (DKIST, NISP), Thirty Meter Telescope (TMT), Giant Magellan Telescope (GMT), Atacama Large Millimeter/submillimeter Array (ALMA), and Europe's EELT project pre-design work (OWL). He established the Jacobus and Michelle Oschmann Scholarship in Optical Sciences and Business Leadership at University of Arizona. Mr. Oschmann received an M.S. in optical sciences and an M.S. in business administration from the University of Arizona.

RACHEL A. OSTEN is a multi-wavelength stellar astronomer at the Space Telescope Science Institute and the deputy mission head for the Hubble Space Telescope. Dr. Osten's research interests include stellar coronae, stellar flares, multi-wavelength observations of flares, stellar radio emission, and flare modeling. Prior to joining the Space Telescope Science Institute, she was a Hubble Fellow at the University of Maryland and NASA Goddard Space Flight Center, and a Jansky Fellow at the National Radio Astronomy Observatory. Dr. Osten is a member of the American Astronomical Society, the International Astronomical Union, and the Atacama Large Millimeter/submillimeter Array (ALMA) North American Science Advisory Committee. She received her Ph.D. from the University of Colorado, Boulder.

LYMAN A. PAGE JR. is the James S. McDonnell Distinguished University Professor in Physics at Princeton University. Dr. Page's primary research is on measurements of the cosmic microwave background (CMB) from ground-based, balloon-borne, and satellite platforms with high-electron mobility transistor (HEMT) amplifiers, superconductor-insulator-superconductor (SIS) mixers, and bolometers. Dr. Page's team first established the existence of a characteristic angular scale in the data, indicating that the universe is spatially flat. He is one of the original co-investigators on the Wilkinson Microwave Anisotropy Probe satellite, whose first-year results provided precision measurements of the universe. Dr. Page was also the founding director of the Atacama Cosmology Telescope project, and is a member of the National Academy of Sciences. He received a Ph.D. in physics from the Massachusetts Institute of Technology. Dr. Page has served on the National Academies' Board on Physics and Astronomy and the Astro2010 Decadal Survey Panel on Radio, Millimeter, and Submillimeter from the Ground.

ELIOT QUATAERT is a professor of astrophysical sciences and Charles A. Young Professor of Astronomy at Princeton University. He was previously a professor of astronomy and physics and the director of the Theoretical Astrophysics Center at the University of California, Berkeley. Dr. Quataert is an astrophysics theorist who works on a wide range of problems, including stars and black holes, plasma astrophysics, and how galaxies form. He has received a number of national awards for his research, including the Warner Prize of the AAS, the Packard Fellowship, a Simons Investigator award from the Simons Foundation, and membership in the American Academy of Arts and Sciences and the National Academy of Sciences. Dr. Quataert received his Ph.D. in astronomy from Harvard University. He has served on the National Academies' Space Studies Board, the Astro2010 Decadal Survey Panel on Stars and Stellar Evolution, the Plasma Science Committee, and the Committee on Plasma 2010: An Assessment of and Outlook for Plasma and Fusion Science.

WANDA A. SIGUR is an independent consultant for both emerging space exploration companies and traditional aerospace industry companies on strategic planning and program management. Ms. Stigur retired from Lockheed Martin as vice president and general manager of the Civil Space business, where she had executive responsibility for national space programs relating to human space flight and space science missions, including planetary, solar, astrophysical, and Earth remote sensing for civil government agencies. These major programs included the Orion Multi-Purpose Crew Vehicle, Hubble and Spitzer space telescopes, GOES-R weather satellites, Juno, GRAIL, MAVEN, Mars Reconnaissance Orbiter, Mars Odyssey, InSight, OSIRIS-REx planetary missions, and the company's nuclear space power programs. She received an M.B.A. from Tulane University. She is a member of the Academy of Medicine, Engineering and Science of Texas (TAMEST) and the National Academy of Engineering. She has served on the National Academies' Space Technology Industry-Government-University Roundtable (STIGUR).

RACHEL SOMERVILLE is a group leader at the Center for Computational Astrophysics at the Flatiron Institute. Dr. Somerville also holds the George A. and Margaret M. Downsbrough Chair in Astrophysics and is a Distinguished Professor at Rutgers University. Her research interests include galaxy formation and evolution, active galactic nuclei, cosmology, and large-scale structure. Dr. Somerville was previously an assistant professor at the University of Michigan, was a senior group leader at the Max Planck Institute for Astronomy, and held a joint appointment at John Hopkins University and the Space Telescope Science Institute. She earned the 2013 Dannie Heinemann Prize for Astrophysics and a 2014 Simons Investigator Award. Dr. Somerville received her Ph.D. in physics from the University of California, Santa Cruz.

KEIVAN G. STASSUN is the Stevenson Endowed Professor of Physics and Astronomy at Vanderbilt University. Dr. Stassun is also the founding director of the Vanderbilt Initiative in Data-Intensive Astrophysics (VIDA). His research focuses on the formation of stars and planetary systems, which increasingly involves approaches at the interface of astronomy, physics, computer science, and informatics. Dr. Stassun currently serves as the general councilor of the American Physical Society and served for 8 years as chair of the American Astronomical Society's Committee on the Status of Minorities. He is known for his leadership and distinction as a scientist and as an innovator in broadening the participation of underrepresented minorities in STEM fields. Dr. Stassun received the 2018 Presidential Award for Excellence in Science, Mathematics, and Engineering Mentoring. He earned a Ph.D. in astronomy from the University of Wisconsin, Madison.

JEAN L. TURNER is professor of physics and astronomy at the University of California, Los Angeles (UCLA). Dr. Turner's research interests include studying gaseous environments of young super star clusters in local galaxies. Prior to joining UCLA, she worked at the Harvard-Smithsonian Center for Astrophysics, and was a visiting scientist at the California Institute of Technology, the Space Telescope Science Institute, and the Joint ALMA Observatory. Dr. Turner is a fellow of the American Association for the Advancement of Science. She received her Ph.D. in astronomy from the University of California, Berkeley. Dr. Turner served on the National Academies' Astro2010 Panel on Radio, Millimeter, and Submillimeter from the Ground, and the Committee on Astronomy and Astrophysics.

PIETER VAN DOKKUM is the Sol Goldman Professor of Astronomy and divisional director of physical sciences and engineering at Yale University. His research interests include stars and stellar populations to the most distant galaxies, along with astronomical instrumentation and telescopes. Prior to joining Yale University, Dr. van Dokkum was a Spitzer Fellow and Hubble Fellow at the California Institute of Technology. He has received numerous awards, including the Marc Aaronson Memorial Prize, the National Science Foundation CAREER Award, and the Pastoor Schmeitz Prize. Dr. van Dokkum received his Ph.D. from the University of Groningen.

ELLEN G. ZWEIBEL is the W.L. Kraushaar Professor of Astronomy and Physics at the University of Wisconsin, Madison. Dr. Zweibel is also the Vilas Distinguished Achievement Professor and past director of the Center for Magnetic Self-Organization. Her research interests and expertise include theoretical astrophysics with a specialty in plasma astrophysics. Prior to joining the University of Wisconsin, Madison, she was a faculty member at the University of Colorado. Dr. Zweibel received numerous awards, including being elected as a fellow of the American Physical Society and the American Physical Society's Maxwell Prize for Plasma Physics, and is a member of the National Academy of Sciences. She received her Ph.D. in astrophysical sciences from Princeton University. Dr. Zweibel has served on the National Academies' Space Studies Board, the Committee on Burning Plasma Assessment, the Panel on Solar Astronomy, and the Committee on Astronomy and Astrophysics.

SCIENCE PANELS

PANEL ON COMPACT OBJECTS AND ENERGETIC PHENOMENA

DEEPTO CHAKRABARTY, *Chair*, is a professor of physics and associate head of the Physics Department at the Massachusetts Institute of Technology (MIT). Dr. Chakrabarty's research interests include observational high-energy astrophysics, neutron stars, accretion disks, and ultracompact stellar binaries. He is a fellow of the American Physical Society and a legacy fellow of the American Astronomical Society, and is also the recipient of several awards and honors, including the Bruno Rossi Prize in High Energy Astrophysics at the American Astronomical Society, the Alfred P. Sloan Research Fellowship, and the Buechner Teaching Prize in Physics at MIT. Dr. Chakrabarty received his Ph.D. in physics from the California Institute of Technology.

LAURA B. CHOMIUK is an associate professor of physics and astronomy at Michigan State University (MSU). After completing her Ph.D., Dr. Chomiuk was a Jansky Fellow of the National Radio Astronomy Observatory at the Smithsonian Astrophysical Observatory. She has far-ranging interests in transient and energetic phenomena, including novae, supernovae, and X-ray binaries, and she pursues these phenomena with multiwavelength observations spanning radio to gamma-ray wavelengths. Dr. Chomiuk is a recipient of a Cottrell Scholarship and a National Science Foundation CAREER Award, along with an MSU Teacher-Scholar Award. She has a Ph.D. in astronomy from the University of Wisconsin, Madison.

DANIEL E. HOLZ is a professor in the Departments of Physics and Astronomy and Astrophysics at the Enrico Fermi Institute and at the Kavli Institute for Cosmological Physics at the University of Chicago. Dr. Holz held postdoctoral appointments at the Albert Einstein Institute (Max Planck Institute for Gravitational Physics), the Kavli Institute for Theoretical Physics in Santa Barbara, the Kavli Institute for Cosmological Physics at the University of Chicago, and as a Richard Feynman Fellow at Los Alamos National Laboratory. His research focuses on general relativity in the context of astrophysics and cosmology, and he is a member of the Laser Interferometer Gravitational-Wave Observatory (LIGO) collaboration. Dr. Holz received a National Science Foundation CAREER Award, a Quantrell Award for Excellence in Undergraduate Teaching, and as a member of LIGO received the Breakthrough Prize in Fundamental Physics. He received a Ph.D. in physics from the University of Chicago.

RAFFAELLA MARGUTTI is an associate professor of physics and astronomy at Northwestern University, and will join the University of California, Berkeley, faculty in fall 2021. Previously, Dr. Margutti was a Harvard postdoctoral fellow and a James Arthur Fellow at New York University. Her research focuses on transient astrophysical phenomena, including stellar explosions, stellar disruptions by supermassive black holes, and compact-object mergers. Her most recent awards include the SLOAN

Fellowship in Physics and the CIFAR Global Scholar Fellowship. Dr. Margutti has a Ph.D. in physics and astronomy from the University of Milano Bicocca.

JULIE MCENERY is a senior scientist of high-energy astrophysics at NASA Goddard Space Flight Center. Dr. McEnery is also the co-director of the Joint Space Science Institute of Goddard and the University of Maryland. She is the senior project scientist for the Nancy Grace Roman Space Telescope and previously served as the project scientist for the Fermi Mission. Dr. McEnery's research focuses on the study of extreme high-energy transients and the development of the ground- and space-based observatories needed to pursue this. She is a fellow of the American Physical Society and a recipient of both the NASA Exceptional Scientific Achievement and the Outstanding Leadership Medals. Dr. McEnery holds a Ph.D. in physics from the University College Dublin.

PETER I. MÉSZÁROS is the Eberly Chair of Astronomy and Astrophysics and a professor of physics at the Pennsylvania State University. Dr. Mészáros is also the director of the Center for Particle and Gravitational Astrophysics at Penn State. His areas of research involve high-energy astrophysics, cosmology, particle astrophysics, gamma-ray bursts, and neutron stars. For the past two decades, Dr. Mészáros has been primarily interested in theoretical aspects of high-energy neutrino astrophysics and multimessenger astrophysics. Awards and memberships include the American Astronomical Society's Rossi Prize, American Academy of Arts and Sciences, Hungarian Academy of Sciences, and the Einstein Professor of Chinese Academy of Sciences. Dr. Mészáros obtained a Ph.D. in astrophysics from the University of California, Berkeley.

RAMESH NARAYAN is the Thomas Dudley Cabot Professor of the Natural Sciences at the Center for Astrophysics, Harvard and Smithsonian, in the Astronomy Department. Previously, Dr. Narayan was on the faculty at the University of Arizona. He is a broad-spectrum theorist with a particular interest in compact objects. Dr. Narayan's research spans a range of topics in high-energy astrophysics, including both black holes and neutron stars; galactic and extragalactic objects; and electromagnetic bands from radio to gamma rays. He was elected to the National Academy of Sciences in 2013 and to the American Association for the Advancement of Science in 2010; he became a fellow of the World Academy of Sciences in 2015 and was elected a fellow of the Royal Society of London (FRS) in 2006. Dr. Narayan received a Ph.D. in physics from Bangalore University, India.

ELIOT QUATAERT, *see steering committee entry above*.

SCOTT M. RANSOM is a tenured astronomer with the National Radio Astronomy Observatory (NRAO) in Charlottesville, Virginia, where he studies pulsars and gravitational waves. Dr. Ransom is also a research professor with the Astronomy Department at the University of Virginia. He works on a wide variety of projects involving finding, timing, and exploiting pulsars of various types, using data from many different instruments and at energies from radio waves to gamma rays. His main focus is on searching for exotic pulsar systems, such as millisecond pulsars and binaries. Once these pulsars are identified, he uses them as tools to probe a variety of basic physics, including tests of general relativity, the emission (and hopefully soon the direct detection) of gravitational waves (as part of the NANOGrav collaboration, of which he is the current chair), and the physics of matter at supra-nuclear densities. Previously, Dr. Ransom was a postdoctoral fellow at McGill University before joining NRAO as a staff astronomer. He won the American Astronomical Society's Helen B. Warner Prize "for a significant contribution to observational or theoretical astronomy during the five years preceding the award." Dr. Ransom is a fellow of the American Physical Society and has authored or co-authored more than 250 refereed publications including more than 20 in *Nature* and *Science*. He received his Ph.D. in astronomy from Harvard University.

TODD A. THOMPSON is a professor of astronomy at Ohio State University. Dr. Thompson was formerly a Hubble Postdoctoral Fellow at the University of California, Berkeley, and a Lyman Spitzer Jr. Postdoctoral Fellow at Princeton University. His areas of research expertise include the mechanism of core-collapse supernovae, gamma-ray bursts, superluminous supernovae, heavy element nucleosynthesis, magnetars, and wide-field transient surveys; star formation, feedback, galactic winds, cosmic rays, and nonthermal emission from galaxies; and binary systems, compact objects, and few-body dynamics. Dr. Thompson was awarded an Alfred P. Sloan Foundation Fellowship, the Ohio State Alumni Award for Distinguished Teaching, a Simons Foundation Fellowship, and an IBM Einstein Fellowship from the Institute for Advanced Study (IAS), Princeton University. He was recently a visiting junior professor while on sabbatical at the IAS. Dr. Thompson received his Ph.D. in physics from the University of Arizona, Tucson.

PANEL ON COSMOLOGY

DANIEL EISENSTEIN, *Chair*, is a professor and department chair at Harvard University. Dr. Eisenstein's research interests include cosmology and extragalactic astronomy, with a mix of theoretical and observational methods. His dominant focus over the past decade has been on the development of the baryon acoustic oscillation method to measure the cosmic distance scale and study dark energy. Prior to joining Harvard University, he was an astronomy faculty member at the University of Arizona and held postdoctoral positions at the Institute for Advanced Study and the University of Chicago. Dr. Eisenstein has been active in the Sloan Digital Sky Survey (SDSS) since 1998 and served as the Director of SDSS-III. He is a member and former co-spokesperson of the Dark Energy Spectroscopic Instrument collaboration, and he is a member of the James Webb Space Telescope (JWST) Near-Infrared Camera instrument team, the SDSS-IV consortium, and the Euclid consortium. Dr. Eisenstein has served as chair of the National Science Foundation Astronomy Portfolio Review committee, and he has been a member of numerous other scientific collaborations and national committees. He has received the Shaw Prize in Astronomy and was named a Simons Investigator. Dr. Eisenstein received his Ph.D. in physics from Harvard University.

LINDSEY E. BLEEM is an assistant physicist at Argonne National Laboratory (ANL). Dr. Bleem's research interests include using clusters of galaxies to constrain cosmological models. She is currently constructing and exploring the properties of new samples of clusters selected via the Sunyaev-Zel'dovich effect using data from the South Pole Telescope, Dark Energy Survey, and Hubble and Spitzer Space Telescopes. Beyond this work, Dr. Bleem is engaged in efforts to better connect simulations and observations of clusters to prepare for the next-generation optical Large Synoptic Survey Telescope (LSST) and cosmic microwave background surveys. She received the Maria Goeppert Fellowship and Sachs Fellowship from the University of Chicago, and the Director's Fellowship from ANL. Dr. Bleem earned her Ph.D. from the University of Chicago.

MARC P. KAMIONKOWSKI, *see steering committee entry above*.

RACHEL MANDELBAUM is a professor at Carnegie Mellon University (CMU). Dr. Mandelbaum was previously an associate research scholar and visiting associate research scholar for the Department of Astrophysical Sciences at Princeton University, and a Hubble Fellow in astrophysics at the Institute for Advanced Study. She has received the AAS Annie Jump Cannon Prize, the Department of Energy Early Career Award, an Alfred P. Sloan Fellowship, and a Simons Investigator Award, and she was the Falco-DeBenedetti Career Development Professor in Physics at CMU. Dr. Mandelbaum's research interests are predominantly in the areas of observational cosmology and galaxy studies. This work includes the use of weak gravitational lensing and other analysis techniques, with projects that range from development of improved data analysis methods to actual application of such methods to existing data. Dr. Mandelbaum

is using data from the Hyper-SuprimeCam (HSC), and she is working on upcoming surveys including the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), Euclid, and the Nancy Grace Roman Space Telescope. She is serving a two-year term as spokesperson of the LSST Dark Energy Science Collaboration (DESC). Dr. Mandelbaum earned her Ph.D. in physics from Princeton University.

MIGUEL F. MORALES is a professor at the University of Washington in the Department of Physics. Dr. Morales is an observational cosmologist and works primarily on measurements of the Epoch of Reionization (EoR) as the Universe's first stars and galaxies burned away the primordial neutral hydrogen fog approximately 13 billion years ago. His radio cosmology group is recognized as an international leader in developing the bespoke instruments and precision data analysis techniques required to reveal the faint cosmological radio signal. The members of this group are builders of the Murchison Widefield Array in western Australia and the Hydrogen Epoch of Reionization Array in South Africa, and have developed one of the four major EoR analysis pipelines. Dr. Morales received the National Science Foundation's CAREER Award and was included in *Scientists Like Me: Faces of Discovery* and as an emerging scholar in *Diversity Magazine*. Dr. Morales earned a Ph.D. in physics at the University of California, Santa Cruz.

DANIEL M. SCOLNIC is an assistant professor of physics at Duke University. Dr. Scolnic was previously a KICP Fellow and a Hubble Fellow at the University of Chicago. He was selected by the Space Studies Board to participate in the National Academy of Sciences and Chinese Academy of Sciences 7th and 8th Forum for New Leaders in Space Science. Dr. Scolnic received a Hubble Postdoctoral Research Fellowship and a Kavli Institute for Cosmological Physics Fellowship, and was a national finalist for the NASA Famelab Competition. He leads dark energy and Hubble constant cosmological analyses using supernovae for the Panoramic Survey Telescope and Rapid Response System (STARRS); the Dark Energy Survey; the LSST; the Hubble Space Telescope Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (CANDELS); Foundation; Supernovae, H0, for the Equation of State (SH0ES); and WFIRST. He is also participating in the design of future missions to find the optical counterparts of gravitational waves. Dr. Scolnic earned his Ph.D. in astronomy from Johns Hopkins University.

MATIAS ZALDARRIAGA is a professor of astrophysics at the Institute for Advanced Study. Dr. Zaldarriaga was previously a professor of astronomy and physics at Harvard University. His research interests include understanding the earliest instants in the history of the Universe and in developing the necessary tools to interpret observations of the Cosmic Microwave Background (CMB) and of the distribution of matter across cosmic history. Dr. Zaldarriaga developed new statistical probes to infer fundamental properties of the Universe from the CMB and radio observations of cosmic reionization. The current generation of CMB experiments is testing models of inflation and gravity using his seminal work. His work on the physics of non-Gaussianity and inflation provides a framework for studying the early Universe. Dr. Zaldarriaga has received a Hubble Fellowship, a David and Lucile Packard Fellowship, the Helen B. Warner Prize from the American Astronomical Society, a Sloan Fellowship, the Gribov Medal from the European Physical Society, and a MacArthur Fellowship. He earned a Ph.D. in physics at the Massachusetts Institute of Technology.

KATHRYN M. ZUREK is a professor of theoretical physics in the Division of Physics, Mathematics, and Astronomy at the California Institute of Technology. Dr. Zurek was previously a senior scientist at the Lawrence Berkeley National Laboratory, a David Schramm Fellow at the Fermi National Accelerator Laboratory, and a member of the Institute for Advanced Study. She has a wide range of research interests, mostly focused at the boundary of particle physics with astrophysics and cosmology. Dr. Zurek's work spans both studies of new physics signatures at colliders as well as astrophysical searches for dark matter and physics beyond the Standard Model in the neutrino sector. She has recently been active in the study of dark matter, working on theories of dark matter and ways to detect it in the laboratory by dark matter-

nucleus interactions, at colliders through high-energy collisions, and in the galaxy by dark matter self-annihilations. Recently, she has been focused on proposing new ideas to detect hidden-sector dark matter in the laboratory. Dr. Zurek earned a Ph.D. in physics from the University of Washington.

PANEL ON GALAXIES

DANIELA CALZETTI, *Chair*, is a professor and head of the Department of Astronomy at the University of Massachusetts, Amherst. Dr. Calzetti's research interests include understanding star formation on the scales of galaxies, using information provided by a variety of both space-borne (Hubble, Spitzer, Herschel, etc.) and ground-based telescopes, at wavelengths that range from the ultraviolet to the radio. Prior to joining the University of Massachusetts, Amherst, she held various positions at the Space Telescope Science Institute, including ESA Fellow, postdoctoral researcher, assistant astronomer, and associate astronomer. Dr. Calzetti is a member of the American Astronomical Society and the International Astronomical Union. In addition, she is a member of the ERC Panel, NASA STDT for the LUVOIR Surveyor Mission Concept Study, AURA Space Telescope Science Institute Council (STIC), and EUCLID Science Consortium/Co-I of Euclid Science Program: Precision Studies of Galaxy Growth and Cosmology. Dr. Calzetti has received numerous awards, including the Award for Outstanding Accomplishments in Research at the University of Massachusetts, the Clarivate-Reuters World's Most Cited Researchers, the Blaauw Professorship of the University of Groningen, and the Tage Erlander Guest Professorship at the University of Stockholm. She received her Ph.D. in astronomy from the University of Rome.

MICHAEL BOYLAN-KOLCHIN is an associate professor of astronomy at the University of Texas, Austin. Dr. Boylan-Kolchin's research focuses on theoretical astrophysics, including numerical simulations of the formation and evolution of cosmological structure, galaxies, and globular clusters; galaxy dynamics; the nature of dark matter; the epoch of reionization; and near-field cosmology. For this panel, he brings expertise with all associated wavelengths. Prior to joining the University of Texas, Dr. Boylan-Kolchin was a faculty member at the University of Maryland and a postdoctoral scientist at the University of California, Irvine, and the Max-Planck-Institut für Astrophysik. He is the recipient of a National Science Foundation CAREER Award. Dr. Boylan-Kolchin received his Ph.D. in physics from the University of California, Berkeley.

HSIAO-WEN CHEN is a professor of astronomy and astrophysics at the University of Chicago. Dr. Chen's research interests include the formation and evolution of galaxies across cosmic time, chemical enrichment in the intergalactic medium (IGM), and transient phenomena. Prior to joining the University of Chicago, she held a postdoctoral position at Carnegie Observatories and a Hubble Fellowship at Massachusetts Institute of Technology. Dr. Chen has served as the vice president of the International Astronomical Union IGM Commission, chair of the Adler Planetarium Visiting Committee, and chair of the Space Telescope Users' Committee. Her work is related to absorption-line spectroscopy of distant light sources (quasars/gamma-ray burst afterglows) to probe diffuse gas around galaxies, and in combining absorption-line observations with galaxy survey data to understand the recycling of baryonic matter between star-forming regions and dark intergalactic medium. For this panel, Dr. Chen brings expertise with ultraviolet wavelengths and spectroscopy. She received her Ph.D. in astronomy from the State University of New York, Stony Brook.

ANN E. HORNSCHEMEIER is the chief of the X-Ray Astrophysics Laboratory at the NASA Goddard Space Flight Center. Dr. Hornschemeier specializes in studies of X-ray emission from accreting black hole and neutron star binary populations, both in the local universe and at cosmologically interesting distances. She chaired the NuSTAR Starburst and Local Group science working group, carrying out observations of nearby galaxies and coordinating observations with Chandra, XMM-Newton, and Swift.

Dr. Hornschemeier is also involved in future missions, serving as the only U.S. science co-investigator on the Wide-Field Imager on Athena and as a member of the international consortium for the Laser Interferometer Space Antenna mission. For this panel, she brings expertise with X-ray wavelengths. Dr. Hornschemeier won the American Astronomical Society's Annie Jump Cannon Award and NASA awarded her an Early Career Achievement Medal. She has been recognized by the American Physical Society as a fellow, and was awarded the Royal Astronomical Society of Canada Helen Sawyer Hogg lectureship for her studies of X-ray emission from galaxies. Dr. Hornschemeier received her Ph.D. in astronomy and astrophysics from Pennsylvania State University.

SUSAN A. KASSIN is an AURA Associate Astronomer with tenure at the Space Telescope Science Institute. Dr. Kassin is also an associate research scientist at the Johns Hopkins University Department of Physics and Astronomy. Previous positions include postdoctoral researcher at the University of California Observatories/Lick Observatory and NASA Postdoctoral Program Fellow at NASA's Goddard Space Flight Center. Dr. Kassin studies galaxy formation and evolution at low and high redshift. She uses observations at optical and near-infrared wavelengths, in addition to numerical simulations. For this panel, Dr. Kassin brings expertise with optical and near-infrared wavelengths. She received a Ph.D. in astronomy and astrophysics from Ohio State University.

AMANDA A. KEPLEY is an assistant scientist with the North American ALMA Science Center at the National Radio Astronomy Observatory (NRAO). Dr. Kepley previously held postdoctoral positions at NRAO and at the University of Virginia. Her research focuses on investigating the role that gas, dust, and magnetic fields within galaxies play in their evolution, primarily using radio telescopes like Atacama Large Millimeter/Submillimeter Array (ALMA), the Karl Jansky Very Large Array (JVLA), and the Green Bank Telescope (GBT). Dr. Kepley also develops and tests heuristics for automated data reduction pipelines, both for ALMA and for her own research. For this panel, she brings expertise with infrared and radio wavelengths. Dr. Kepley was the recipient of a National Science Foundation Graduate Research Fellowship. She received her Ph.D. in astronomy from the University of Wisconsin, Madison.

CHARLES C. STEIDEL is the Lee A. DuBridge Professor of Astronomy at the California Institute of Technology. Dr. Steidel's previous positions include assistant professor of physics at the Massachusetts Institute of Technology (MIT) and Hubble Fellow at the University of California, Berkeley. He is a world leader in observational cosmology. Dr. Steidel defined the state and chemical composition of the intergalactic medium in the early Universe, and he discovered normal star-forming galaxies at high redshift. With co-workers, he measured the clustering of these galaxies, thus placing serious constraints on cosmological models. Dr. Steidel's expertise is related to the processes of galaxy formation and the nature of the intergalactic medium. He has received the Gruber Cosmology Prize from the Peter and Patricia Gruber Foundation in recognition of his revolutionary studies of the most distant galaxies in the universe. Dr. Steidel received his Ph.D. in astronomy from the California Institute of Technology.

DANIEL K. STERN is the NuSTAR project scientist at NASA's Jet Propulsion Laboratory. Dr. Stern's research interests emphasize understanding the cosmic history of black hole formation and activity, observational cosmology, and identifying and studying galaxies and galaxy clusters at high redshift. For this panel, he brings expertise with X-ray, ultraviolet, visible, and infrared wavelengths. Some of Dr. Stern's recent awards are the NASA Exceptional Scientific Achievement Medal, the NASA Group Achievement Award to the NuSTAR Science Team, and the NASA Exceptional Achievement Medal. He received his Ph.D. in astrophysics from the University of California, Berkeley.

TOMMASO TREU is a professor at the University of California, Los Angeles (UCLA), in the Division of Astronomy and Astrophysics. Previous positions include Distinguished Visitor at the Space Telescope Science Institute and professor at the University of California, Santa Barbara (UCSB). Dr. Treu's research interests and expertise include galaxy formation and evolution. In particular, he is interested in early-type

galaxies; galaxies in clusters; high redshift galaxies; the co-evolution of spheroids and black holes; gravitational lensing and dark matter in galaxies; and clusters of galaxies, cosmography from gravitational time delays, and galaxies in the epoch of reionization. For this panel, Dr. Treu brings expertise with optical and infrared wavelengths. He is a recipient of the American Astronomical Society Newton Lacy Pierce Prize, the UCSB H.J. Plous Memorial Award, and the David and Lucille Packard Fellowship for Science and Engineering. Dr. Treu received his Ph.D. in physics from Scuola Normale Superiore, Pisa, Italy.

PIETER VAN DOKKUM, *see steering committee entry above*.

DAVID H. WEINBERG is a Distinguished University Professor and chair of the Department of Astronomy at Ohio State University. Dr. Weinberg studies the large-scale structure of the universe, dark energy and dark matter, the formation and evolution of galaxies and quasars, and the intergalactic medium (IGM). He is well-known for his development of “halo occupation” methods to connect observed galaxy clustering to underlying dark matter structure, for theoretical modeling and cosmological applications of the Lyman-alpha forest, and for numerical simulation studies of the mechanisms of galaxy growth. For this panel, Dr. Weinberg brings expertise with all associated wavelengths. He has received the University Distinguished Scholar award and the Lancelot M. Berkeley New York Community Trust Prize for Meritorious Work in Astronomy by the American Astronomical Society. Dr. Weinberg received his Ph.D. in astrophysical sciences from Princeton University.

PANEL ON EXOPLANETS, ASTROBIOLOGY, AND THE SOLAR SYSTEM

VICTORIA S. MEADOWS, *Chair*, is a professor of astronomy at the University of Washington in the Department of Astronomy, where she is also director of the Astrobiology Program and principal investigator for the NASA Virtual Planetary Laboratory. Dr. Meadows’s research interests include theoretical modeling of terrestrial planetary environments to understand their habitability, the generation and detectability of exoplanetary biosignatures and their false positives, and solar system planetary observations. The overarching goal of her research is to determine how to recognize whether a distant extrasolar planet can or does support life. Previously, Dr. Meadows was a research scientist at the Jet Propulsion Laboratory and an associate research scientist at the Spitzer Science Center at the California Institute of Technology. She is a recipient of several NASA Group Achievement Awards, has been on the SETI Institute Science Advisory Board, and was a Frontiers of Science Kavli Fellow. Dr. Meadows earned her Ph.D. in physics from the University of Sydney.

DAVID A. BRAIN is an associate chair for undergraduate studies at the University of Colorado (CU). Dr. Brain is also an associate professor of astrophysical and planetary sciences in the Department of Astrophysical and Planetary Sciences and the Laboratory for Atmospheric and Space Physics. At CU, he is a co-deputy principal investigator of NASA’s MAVEN spacecraft mission at Mars, and project scientist and advisor for the United Arab Emirates Hope spacecraft mission to Mars. Dr. Brain’s research interests include atmospheric escape and long-term evolution of planetary atmospheres, planetary magnetospheres and plasma interactions, and the influence of planetary magnetic fields on climate evolution and habitability. Previously, he was a research physicist at the Space Sciences Laboratory at the University of California, Berkeley. Dr. Brain is a recipient of the NASA Early Career Fellowship in planetary sciences. He earned his Ph.D. in astrophysical and planetary sciences from the University of Colorado.

IAN J.M. CROSSFIELD is an assistant professor of astrophysics at the Massachusetts Institute of Technology (MIT) in the Department of Physics. At MIT, Dr. Crossfield has led the discovery and characterization of new exoplanets discovered by NASA’s TESS and Kepler/K2 missions. In this effort, he leads a number of large observational programs with the Hubble Space Telescope (transmission

spectroscopy), Spitzer Space Telescope (transit and secondary eclipse photometry), 10 m Keck Observatory (precise radial velocities), and 8.2 m Gemini Observatory (diffraction-limited adaptive optics and speckle imaging). His research interests focus on the characterization of exoplanet atmospheres to test models of planet formation and of atmospheric chemistry, thermal structure, and general circulation. Previously, Dr. Crossfield was a NASA Sagan Postdoctoral Fellow at the University of California, Santa Cruz, in the Department of Astronomy and at the University of Arizona's Lunar and Planetary Lab, and a postdoctoral fellow at the Max Planck Institute for Astronomy in Heidelberg, Germany. He also was a systems engineer at NASA's Jet Propulsion Laboratory for 3 years, after which he earned his Ph.D. in astrophysics from the University of California, Los Angeles.

COURTNEY D. DRESSING is an assistant professor at the University of California, Berkeley, in the Department of Astronomy. Dr. Dressing is an observational astronomer focused on detecting and characterizing planetary systems. She conducts both statistical investigations of the ensemble of known planetary systems and in-depth studies of individual systems. Dr. Dressing's research group uses telescopes on the ground and in space to search for planets, determine their orbital parameters, measure their masses, and constrain their bulk compositions. She is curious about planet formation and evolution, the frequency of planetary systems in the Galaxy, and the prospects for detecting life on planets outside our Solar System. Previously, Dr. Dressing was a NASA Sagan Fellow at the California Institute of Technology. She was awarded a Sloan Research Fellowship in 2019 for becoming "a world leader in the search for other worlds." In 2019, Dr. Dressing was also awarded a Hellman Fellowship and a Packard Fellowship from the David and Lucile Packard Foundation. She earned a Ph.D. in astronomy and astrophysics from Harvard University.

JONATHAN J. FORTNEY, *see steering committee entry above*.

TIFFANY KATARIA is a scientist at the Jet Propulsion Laboratory in the Astrophysics and Space Sciences Division. Dr. Kataria's research focuses on the theoretical modeling of dynamics and chemistry in the atmospheres of transiting and directly imaged exoplanets, and particularly how theoretical models can be used to interpret observations of exoplanet atmospheres using ground- and space-based telescopes. She is currently a member of the executive committee of the Exoplanetary Program Analysis Group (ExoPAG) and a member of the JWST Users Committee (JSTUC). Prior to joining the Jet Propulsion Laboratory, Dr. Kataria was a postdoctoral fellow at the University of Exeter. She received her Ph.D. in planetary sciences from the University of Arizona.

KATHLEEN E. MANDT is the chief scientist for exoplanets at the Johns Hopkins University (JHU) Applied Physics Laboratory (APL). At APL, Dr. Mandt serves as the chief scientist for exoplanets, where she is responsible for initiating an exoplanet program that leverages the planetary science, heliophysics, mission leadership, and instrument development expertise at APL to contribute to future detection and characterization of exoplanets. She serves in several community and NASA mission leadership roles, including membership on the Outer Planets Assessment Group (OPAG) steering committee, and served on the Division for Planetary Science Professional Culture and Climate Subcommittee. Dr. Mandt was the Volatiles Theme Lead for the Lunar Reconnaissance Orbiter (LRO) mission and is the project scientist for the LRO Lyman Alpha Mapping Project (LAMP) instrument, the project scientist for the Phase A Io Volcano Observer (IVO) Discovery Mission concept study, the deputy project scientist for the Heliophysics Division-funded Interstellar Probe predecadal mission study, and a science team member on the Europa Clipper Plasma Instrument for Magnetic Sounding teams. Dr. Mandt was a science team member on the Cassini Ion Neutral Mass Spectrometer and Rosetta Ion Electron Spectrometer teams. Her research covers a broad range of topics, including the dynamics, chemistry, and evolution of planetary atmospheres. Dr. Mandt is particularly interested in leveraging the expertise of the planetary science community to advance characterization of exoplanet atmospheres and applying studies in solar system atmospheric evolution to better understand the evolution of exoplanet systems. Previously, Mandt was an

adjunct professor in the Department of Physics and Astronomy at the University of Texas, San Antonio, and a senior research scientist at Southwest Research Institute. She earned her Ph.D. in environmental science and engineering from the University of Texas, San Antonio.

MARK S. MARLEY is a research scientist at NASA Ames Research Center. At NASA Ames, Dr. Marley primarily studies the atmospheres of extrasolar giant planets and brown dwarfs through theoretical modeling and comparisons to data. His research interests include the chemistry and physics of clouds and hazes, departures from chemical equilibrium, the origin and evolution of extrasolar giant planets, the characterization of extrasolar planets through direct imaging, giant planet seismology, and synergies between Solar System and extrasolar planetary science. Previously, Dr. Marley was on the faculty of New Mexico State University in the Department of Astronomy. He is a fellow of the American Astronomical Society and an associate fellow of Ames Research Center, and has twice been awarded the NASA Medal for Exceptional Scientific Achievement. Dr. Marley earned his Ph.D. in planetary science from the University of Arizona.

BRITNEY E. SCHMIDT is an associate professor at the Georgia Institute of Technology. Dr. Schmidt is the principal investigator of the Ross Ice Shelf and Europa Underwater Probe (RISE-UP), an interdisciplinary astrobiology and oceanographic investigation leveraging remote sensing and autonomous underwater vehicles to examine Earth’s ice shelves as analogs for extraterrestrial icy moons and their potential for habitability. Her research interest in the astrobiology of icy systems focuses on Europa, where she models the formation of surface terrain to better understand ice-ocean interactions and works on a variety of instrument technology and platforms for subsurface exploration. Dr. Schmidt is also a co-investigator on NASA’s Europa Clipper radar team, a member of the Europa Lander and LUVOIR science definitions teams, and an associate of the Dawn mission. She was previously a postdoctoral fellow at the University of Texas, Austin, where she was named outstanding early career researcher, and she is recipient of a NASA Early Career Fellowship and the Eric R. Immel Memorial Award for Excellence in Teaching from the Georgia Tech College of Science. Dr. Schmidt earned her Ph.D. in geophysics and space physics from the University of California, Los Angeles.

CHRISTOPHER C. STARK is an associate scientist at the Space Telescope Science Institute (STScI). Dr. Stark has led the exoplanet science yield simulations for many of the direct-imaging mission concepts currently under study and is a leading expert in exozodiacal dust and debris disks. His research interests include debris disks/exozodis (as a source of both signal and problematic noise), disk composition, planet-dust dynamics and gravitationally induced disk structures, high-contrast direct-imaging methods and instrument design (including coronagraphy, external occultation, and interferometric nulling), the optimization of observations to maximize the scientific return of direct-imaging missions, and systems design for future exoplanet-imaging missions. Prior to working at STScI, Dr. Stark was a NASA Postdoctoral Program Fellow at NASA’s Goddard Space Flight Center and a Carnegie Postdoctoral Fellow at the Carnegie Department of Terrestrial Magnetism. He earned his Ph.D. in physics from the University of Maryland.

PANEL ON THE INTERSTELLAR MEDIUM AND STAR AND PLANET FORMATION

LEE W. HARTMANN, *Chair*, is the Leo Goldberg Collegiate Professor of Astronomy at the University of Michigan. Dr. Hartmann has worked as an astrophysicist at the Smithsonian Astrophysical Observatory and was a vice president of the American Astronomical Society. His research interests include the formation of stars and star clusters, molecular cloud structure and dynamics, protostellar accretion, evolution of protoplanetary disks and planet formation, and mass function of stars. Dr. Hartmann is a fellow of the American Association for the Advancement of Science. He received his Ph.D. in astronomy from the University of Wisconsin System.

SEAN M. ANDREWS is an astrophysicist with the Smithsonian Astrophysical Observatory (SAO) and a lecturer on astronomy at Harvard University at the Center for Astrophysics, Harvard and Smithsonian. Prior to that, Dr. Andrews was a Hubble Fellow at SAO. He does research on planet formation in the disks of gas and dust that orbit around young stars, primarily using radio interferometer data. In 2010 and 2017, Dr. Andrews was awarded the Secretary's Research Prize by the Smithsonian Institution. He was the principal investigator of the Disk Substructures at High Angular Resolution Project (DSHARP), an Atacama Large Millimeter/Submillimeter Array (ALMA) Large Program. Dr. Andrews received his Ph.D. in astronomy from the University of Hawaii.

PHILIP J. ARMITAGE is a professor at Stony Brook University in the Department of Physics and Astronomy. Dr. Armitage is also a group leader at the Center for Computational Astrophysics at the Flatiron Institute. Prior to his current position, he did postdoctoral work at the University of Toronto's Canadian Institute for Theoretical Astrophysics and at the Max Planck Institute for Astrophysics in Munich, Germany. Dr. Armitage is interested in understanding the physical processes involved in the formation of planetary systems, including the role of two-fluid instabilities in forming the first analogs of asteroids and comets, and the importance of radiation hydrodynamic effects in the accretion of planetary envelopes. He received his Ph.D. from the University of Cambridge, Institute of Astronomy.

BRUCE T. DRAINE is a professor of astrophysical sciences at Princeton University. After serving for 2 years in the Peace Corps, teaching physics and math in Ghana, Dr. Draine went to Cornell University. He has computed models for interstellar dust-grain properties, proposed a solution to the problem of the polarization of starlight by dust, showed that emissions from spinning grains is an important foreground to the cosmic background radiation, and discovered a new kind of interstellar shock wave. His work has created important models to interpret some of the processes that occur in interstellar space. Dr. Draine was a member of the Spitzer Infrared Nearby Galaxy Survey (SINGS), a Spitzer Space Telescope Legacy project that studied a sample of 75 nearby galaxies, and a member of the KINGFISH collaboration, a Herschel Key Project using the Herschel Space Telescope (a 3.5 m telescope in space with a cryogenic focal plane) to study 61 nearby ($d < 30$ Mpc) galaxies using both far-infrared imaging and spectroscopy. Dr. Draine received his Ph.D. in theoretical physics from Cornell University.

KAITLIN M. KRATTER is an associate professor of astronomy at the University of Arizona. Prior appointments include a Hubble Fellowship at the University of Colorado and an Institute for Theory and Computation Fellowship at Harvard University. Dr. Kratter employs analytic and computational techniques to tackle topics including accretion disk dynamics, binary formation, few-body dynamics, and planet-disk interactions. She received her Ph.D. in astrophysics from the University of Toronto.

KARIN M. SANDSTROM is an assistant professor of astrophysics at the University of California, San Diego. Dr. Sandstrom specializes in multiwavelength studies of the interstellar medium (ISM) in nearby galaxies. Her interests include interstellar dust, ISM phases, heating and cooling of gas and dust, feedback from stellar populations, and chemical enrichment. Dr. Sandstrom has been awarded observing time as principal investigator on numerous telescopes, including the Atacama Large Millimeter Array (ALMA), the Hubble Space Telescope (HST), and the Karl Jansky Very Large Array (JVLA). She received her Ph.D. from the University of California, Berkeley.

SNEZANA STANIMIROVIC is a professor of astronomy at the University of Wisconsin, Madison. Dr. Stanimirovic moved to Madison after working as a research associate in the Radio Astronomy Laboratory at the University of California, Berkeley. Before Berkeley, she spent 3 years in Puerto Rico, working at the Arecibo Observatory. Dr. Stanimirovic's research interests include mapping neutral hydrogen in and around the Milky Way; statistical investigation of the interstellar medium; and star formation, magnetic fields, and diffuse matter in the galaxy. She was awarded the NSF Career Award in 2010 and became a

fellow of the American Association for the Advancement of Science and a Guggenheim Fellow. Dr. Stanimirovic earned her Ph.D. at the University of Western Sydney, jointly supervised by the Australia Telescope National Facilities.

ELLEN G. ZWEIBEL, *see steering committee entry above*.

PANEL ON STARS, THE SUN, AND STELLAR POPULATIONS

SARBANI BASU, *Chair*, is a professor and chair of the Department of Astronomy at Yale University. Dr. Basu's research interests include the study of the Sun and other stars using data on stellar oscillations, and in studying the variations in the Sun over time scales that are of societal relevance. She is a co-investigator of the Helioseismic and Magnetic Imager on NASA's Solar Dynamics Observatory, and a member of the steering committee of the TESS Asteroseismic Science Consortium. Dr. Basu serves as the deputy chair of the board of directors of the Association of Universities for Research in Astronomy (AURA), and she is a member of the advisory board of the journal *Solar Physics*. She is the recipient of numerous awards: she is a 2020 fellow of the American Astronomical Society, received the 2018 George Ellery Hale Prize of the Solar Physics Division of the American Astronomical Society, is a 2015 fellow of the American Association for the Advancement of Science, and received the 1996 M.K. Vainu Bappu Gold Medal of the Astronomical Society of India. Dr. Basu received her Ph.D. in physics from Tata Institute of Fundamental Research.

NANCY S. BRICKHOUSE is the senior science advisor at the Center for Astrophysics, Harvard and Smithsonian. Dr. Brickhouse has served as the associate director for the Solar, Stellar, and Planetary Sciences Division at the Center for Astrophysics. Her research interests include solar and stellar coronal physics, plasma spectral modeling, atomic data for astrophysics, ultraviolet to X-ray spectroscopy of diverse objects, and physical processes in astrophysical plasmas. Dr. Brickhouse is a leader of the Atomic Data for Astrophysicists (ATOMDB) Project, which uses collisional and radiative atomic data to generate spectral models needed for high-energy astrophysics. She received her Ph.D. in physics from the University of Wisconsin, Madison.

ADAM BURGASSER is a professor in the Department of Physics at the University of California, San Diego, and an observational astrophysicist whose research interests include the lowest mass stars, brown dwarfs, and extrasolar planets. Dr. Burgasser's work focuses on substellar atmospheres, multiple systems, activity, and populations using-optical/infrared spectroscopy, high-resolution imaging, radio interferometry, and large data science. He also conducts research in physics education and art-science collaboratories. Dr. Burgasser has previously held a Hubble Postdoctoral Fellowship at the University of California, Los Angeles; a Spitzer Postdoctoral Fellowship at the American Museum of Natural History; and a faculty position in physics at the Massachusetts Institute of Technology. He has been awarded the University of California, San Diego's Equal Opportunity/Affirmative Action and Diversity Award, Outstanding Mentor Award, and Distinguished Teaching Award, and was a faculty Fulbright Scholar at the University of Exeter. Dr. Burgasser is a member of the International Astronomical Union, National Society of Black Physicists, and SACNAS, and is a vice president of the American Astronomical Society. He received his Ph.D. in physics from the California Institute of Technology.

JULIANNE DALCANTON, *see steering committee entry above*.

JENNIFER A. JOHNSON is a professor of astronomy at Ohio State University (OSU). Dr. Johnson's research interests include stellar abundances, origin of the elements, nucleocosmochronology, and the formation of the Galaxy and the local group. She is the program head of the Milky Way Mapper of the Sloan Digital Sky Survey. Previously, Dr. Johnson was a postdoctoral fellow at the Carnegie Institution

for Science and at the Herzberg Institute of Astrophysics. She received her Ph.D. in astronomy from the University of California, Santa Cruz.

R.T. JAMES MCATEER is an associate professor at New Mexico State University (NMSU). Dr. McAteer also serves as director of the Sunspot Solar Observatory. His research interests are in space weather monitoring and solar cycle studies, understanding the physics of solar flares and coronal mass ejections, and the heating of the solar atmosphere. Dr. McAteer is the principal investigator of the Solar Physics and Space Weather research group at NMSU, where he leads a convergence program of computer science, electrical engineering, and astrophysics to facilitate interdisciplinary research and space weather predictions. He chairs the National Solar Observatory Users Committee and is a current member of the Daniel K. Inouye Data Policy Advisory Committee. Prior to joining NMSU, Dr. McAteer was an EU Marie Curie Research Fellow at Trinity College Dublin, a STEREO senior scientist and research associate at NASA Goddard Space Flight Center, and a Leverhulme Trust Research Fellow at Queen's University Belfast. He is a recipient of the NSF Faculty CAREER Award. Dr. McAteer received his Ph.D. in physics from the Queen's University Belfast.

ELISA V. QUINTANA is an astrophysicist at the NASA Goddard Space Flight Center. Dr. Quintana serves as the TESS deputy project scientist, the Nancy Grace Roman Space Telescope deputy project scientist for communications, and the principal investigator of the Pandora SmallSat mission to study exoplanet atmospheres. She is also a member of the TESS Guest Investigator Office and a member of the Goddard Diversity and Inclusion Advisory Committee. Dr. Quintana's research is focused on the detection, characterization, and formation of exoplanets using ground- and space-based observations, data analysis techniques, and modeling. She leads the Goddard Exoplanet Group, which works on research related to exoplanets, low-mass star activity, and developing small space-based mission concepts. Previously, Dr. Quintana worked at NASA Ames Research Center and the SETI Institute for 10 years on the Kepler and K2 space missions. She led a team of astronomers to confirm Kepler-186f, the first Earth-size planet found to orbit within the habitable zone of another star. Dr. Quintana is the recipient of the Great Minds in STEM 2015 Scientist of the Year, the Lupe Ontiveros Dream Award, and the NASA Software of the Year Award. She is a member of the American Astronomical Society. Dr. Quintana received her Ph.D. in physics from the University of Michigan, Ann Arbor.

LOUIS-GREGORY STROLGER is the AURA observatory scientist at the Space Telescope Science Institute. Dr. Strolger is also an associate research scientist in the Department of Physics and Astronomy at Johns Hopkins University. His interests include supernovae, supernova cosmology, and dark energy. Dr. Strolger is primarily interested in the nature of supernovae progenitors through bulk analyses of rates and environmental effects (e.g., star-formation, metallicity) and the evolution of these properties over cosmic history. Prior to this, he was an associate professor of physics and astronomy at Western Kentucky University. Dr. Strolger received his Ph.D. in astronomy and astrophysics from the University of Michigan, Ann Arbor.

PROGRAM PANELS

PANEL ON AN ENABLING FOUNDATION FOR RESEARCH

DAVID N. SPERGEL, *Chair*, is the founding director for computational astrophysics at the Flatiron Institute and the Charles A. Young Professor of Astronomy emeritus of the Department of Astrophysical Sciences at Princeton University. He is currently co-chair of the WFIRST-AFTA science team and the editor of the *Princeton Series in Astrophysics*. Dr. Spergel has made major contributions to cosmology, astroparticle physics, galactic structure, and instrumentation. He led the theoretical analysis for the

Wilkinson Microwave Anisotropy Probe (WMAP) and for the Atacama Cosmology Telescope, invented novel coronagraphs for planet detection, originated and explored the concept of self-interacting dark matter, and showed that the Milky Way is a barred galaxy. He was the W.M. Keck Distinguished Visiting professor of Astrophysics at the Institute for Advanced Study, and an Alfred P. Sloan Research fellow. Dr. Spergel has received numerous awards, including the NSF Presidential Young Investigator Award, the Helen B. Warner Prize, the Bart Bok Prize, the AAS Second Century Lecturer, a MacArthur Fellowship, the Shaw Prize in Astrophysics, the Breakthrough Prize in Fundamental Physics, and shared the Gruber Prize as a member of the Wilkinson Microwave Anisotropy Probe (WMAP) science team. He is a fellow of the American Physical Society, an honorary member of the National Society of Black Physicists, and a member of the National Academy of Sciences. Dr. Spergel served on the NSF's Advisory Committee for Astronomical Sciences; the Theory, Experimental and Laboratory Astrophysics Subcommittee; and the Scientific Advisory Board for the Hayden Planetarium. He received his Ph.D. in astronomy from Harvard University.

MICHAEL BLANTON is a professor of physics at New York University. Dr. Blanton specializes in computation and modeling of large-scale structures in the universe and is currently director of the Sloan Digital Sky Survey IV and was Data Coordinator of Sloan Digital Sky Survey III. He received his Ph.D. in astrophysics from Princeton University.

KELLE L. CRUZ is an associate professor of astrophysics and astronomy at Hunter College, City University of New York. Dr. Cruz's research interests are observational study of low mass stars and brown dwarfs, optical and near-infrared spectroscopy, cool atmospheres, and stellar content of the Solar Neighborhood. Much of her research concentrates on creating large public data sets of very low-mass stars and brown dwarfs and using that data to undertake statistically robust studies of their physical properties. Dr. Cruz is also the founder and editor for AstroBetter, a blog and wiki site for professional astronomers. She was a National Science Foundation Astronomy and Astrophysics Postdoctoral Fellow at the American Museum of Natural History in New York City and then a Spitzer Fellow at Caltech. She received her Ph.D. in physics and astronomy from the University of Pennsylvania.

MARK J. DEVLIN is the Reese W. Flower Professor of Astronomy and Astrophysics at the University of Pennsylvania. Dr. Devlin specializes in experimental cosmology at millimeter and submillimeter wavelengths, collecting data from which he makes statistical inferences about the evolutionary history of the universe. He designs and builds instrumentation and telescopes that he uses to observe from high-altitude balloons and the high plateaus of Chile. Dr. Devlin received the University of Pennsylvania School of Arts and Sciences Ira H. Abrams Memorial Award for Distinguished Teaching and was an Alfred P. Sloan Fellow. He received his Ph.D. in physics from the University of California, Berkeley.

MEGAN E. DONAHUE is professor of astronomy at Michigan State University. She studies galaxy clusters, including models and observational tests of cooling flows in the gas within clusters. Dr. Donahue is a fellow of the American Physical Society and is currently the president of the American Astronomical Society. She earned her Ph.D. in astrophysics from the University of Colorado.

KEITH A. HAWKINS is an assistant professor of astronomy at the University of Texas, Austin. Dr. Hawkin's research interests are galactic and stellar archaeology, chemical composition of stars, stellar spectroscopy, exoplanet host star characterization, and galactic structure. He began as an assistant professor of astronomy at the University of Texas, where he was named a Scialog Fellow and a Kavli Fellow. He is a member of the American Astronomical Society and serves on their Committee for the Status of Minorities in Astronomy (CSMA), and he was chair, King's College Graduate Society. Dr. Hawkins was a British Marshall Scholar and has received a Simons Foundation Junior Research Fellowship at Columbia University. He received his Ph.D. in astronomy from the University of Cambridge.

ALINA A. KIESSLING is a research scientist at the Jet Propulsion Laboratory. Dr. Kiessling has a background in dark matter and dark energy research through weak lensing analysis of N-body simulations. She is currently working on the upcoming Euclid Space Telescope, the Vera C. Rubin Observatory, and the Nancy Grace Roman Space Telescope. Formerly, she co-led an investigation on stratospheric airships, with the goal of determining whether these may become low-cost platforms for astrophysics (and Earth science) missions in the future. Dr. Kiessling received her Ph.D. from the University of Edinburgh.

KARIN ÖBERG is a professor of astronomy and director of Undergraduate Studies at the Center for Astrophysics, Harvard and Smithsonian. Dr. Öberg is also leader of the Öberg Astrochemistry Group at the Center and specializes in astrochemistry and its impact on planet formation, including the compositions of nascent planets. She is a recipient of the Hubble Postdoctoral Fellowship, Alfred P. Sloan Fellowship in Physics, and the Packard Fellowship for Science and Engineering. She received her Ph.D. from Leiden University.

ANGELA V. OLINTO is the dean of the Division of the Physical Sciences at the University of Chicago. Dr. Olinto is also the Albert A. Michelson Distinguished Service Professor in the Department of Astronomy and Astrophysics and the Kavli Institute for Cosmological Physics at the University of Chicago. She previously served as chair of the Department of Astronomy and Astrophysics. Dr. Olinto is best known for her contributions to the study of the structure of neutron stars, primordial inflationary theory, cosmic magnetic fields, the nature of the dark matter, and the origin of the highest energy cosmic rays, gamma-rays, and neutrinos. She is the principal investigator of the Probe of Extreme Multi-Messenger Astrophysics (POEMMA) space mission, the principal investigator of the Extreme Universe Space Observatory (EUSO) on a super pressure balloon mission, and was a member of the Pierre Auger Observatory. She is a member of the National Academy of Sciences and of the American Academy of Arts and Sciences. She is a fellow of the American Physical Society and the American Association for the Advancement of Science. She received the Chaire d'Excellence Award of the French Agence Nationale de Recherche, the Llewellyn John and Harriet Manchester Quantrell Award for Excellence in Undergraduate Teaching, and the Faculty Award for Excellence in Graduate Teaching at the University of Chicago. Dr. Olinto received her Ph.D. in physics at the Massachusetts Institute of Technology and her B.S. in physics at the Pontificia Universidade Catolica of Rio de Janeiro, Brazil.

BERNARD J. RAUSCHER is an experimental astrophysicist at NASA Goddard Space Flight Center. He serves as a detector scientist within the James Webb Space Telescope (JWST) Project and as a member of the JWST Near Infrared Spectrograph (NIRSpec) Science Team. Dr. Rauscher has experience in detector development, including near-infrared detector arrays, photon counting CCDs, and most recently superconducting single photon detectors. His work has led to the development of Improved Reference Sampling and Subtraction for JWST NIRSpec and new algorithms for testing space flight hardware. He received his Ph.D. from the University of Chicago.

RACHEL SOMERVILLE, *see steering committee entry above*.

JAMES M. STONE is a professor at the Institute for Advanced Study in the School of Natural Sciences. Previously, Dr. Stone was the Lyman Spitzer, Jr. Professor of Astrophysical Sciences and a professor of applied and computational mathematics at Princeton University as well as the chair of the university's Department of Astrophysical Sciences. His research interests include star formation, accretion flows, interstellar gas dynamics, and the development of numerical algorithms for magnetohydrodynamics and radiation hydrodynamics. The public codes ZEUS-2D, released in 1992 by Stone and Michael Norman, and Athena, released in 2008 by Stone and his collaborators, are among the most powerful and widely used codes for astrophysical fluid dynamics. Dr. Stone was named a fellow of the American Physical

Society and received the organization's Aneesur Rahman Prize for Computational Physics and the Dirk Brouwer Career Award from the American Astronomical Society. During his academic career, Dr. Stone has held academic positions at the Princeton University, the University of Cambridge, and the University of Maryland. He is also member of the American Astronomical Association, the American Physical Society, and the International Astronomical Union, and the American Academy of Arts and Science. Stone received his Ph.D. for astronomy from the University of Illinois. He has previously served on several Academies' committee.

PANEL ON ELECTROMAGNETIC OBSERVATIONS FROM SPACE 1

MARCIA J. RIEKE, *Chair*, is a Regents Professor of Astronomy and an astronomer at the University of Arizona in the Department of Astronomy and Steward Observatory. Her research interests include infrared observations of galactic nuclei and high-redshift galaxies. Dr. Rieke has served as the deputy principal investigator on the near-infrared camera and multi-object spectrometer for HST (NICMOS), and she is currently the principal investigator for the near-infrared camera (NIRCam) for the James Webb Space Telescope. She has worked on the Spitzer Space Telescope as a co-investigator for the multiband imaging photometer and as an outreach coordinator and as a member of the Science Working Group. Rieke was also involved with several infrared ground observatories, including the Multiple Mirror Telescope in Arizona. She is a fellow of the American Academy of Arts and Sciences and a member of the National Academy of Sciences. Dr. Rieke received her Ph.D. in physics from the Massachusetts Institute of Technology.

RUSLAN BELIKOV is an astrophysicist at NASA Ames Research Center. He is also director of the Exoplanet Technologies research group at NASA Ames, which has demonstrated several state-of-the-art milestones in high-contrast imaging. In addition, Dr. Belikov and his team have been pioneering and advancing technologies to suppress starlight in multi-star systems such as Alpha Centauri to enable direct imaging of exoplanets there. Dr. Belikov has served on NASA's Exoplanet Program Analysis Group executive committee, where he chaired a Science Analysis Group to survey exoplanet statistics. He has more than a decade of experience in developing technologies and mission concepts to directly image exoplanets, especially potentially habitable ones. Dr. Belikov has a Ph.D. in electrical engineering from Stanford University.

REBECCA A. BERNSTEIN is a staff scientist at the Carnegie Science Institute. Dr. Bernstein's research has focused on measurements of the diffuse extragalactic backgrounds at optical wavelengths. That work led to an interest in the technical aspects of low surface brightness measurements, stellar spectroscopy, and instrument design. As a Hubble Fellow at the Carnegie Observatories, Dr. Bernstein designed and led the development of the echelle spectrograph for the Magellan telescopes (MIKE, commissioned 2001) while developing a method for measuring the detailed chemical abundances of unresolved, extragalactic globular clusters. While on the faculty at the University of Michigan (UM), she also designed the optics for the Folded-port InfraRed Echellette (FIRE) spectrograph for the Magellan telescopes and the prime focus Dark Energy Survey Camera (DECam) used for the DES survey at the CTIO's 4m Blanco telescope in Chile. After earning tenure at UM, Bernstein moved UC Santa Cruz, where she was the principal investigator and optical designer for the wide-field optical spectrograph for Thirty Meter Telescope (TMT), one of its planned first-light instruments. Bernstein served as staff astronomer at the Carnegie Observatories and became the project scientist for the Giant Magellan Telescope (GMT). She earned her Ph.D. in astrophysics from the California Institute of Technology.

LESTER M. COHEN is the retired chief engineer of the Structural Analysis and Design Group at the Center for Astrophysics: Harvard & Smithsonian (CfA). Mr. Cohen's areas of expertise include structures, structural mechanics, and mounting and fabrication of optics. He served as lead mechanical

engineer, NASA JWST Optical Telescope Element. Mr. Cohen has earned 12 NASA group and individual awards, including two NASA Public Service Medals and one Distinguished Public Service Medal for his work on two of NASA's Great Observatories: Chandra and JWST. He has an M.S. in civil engineering from Northeastern University.

NIKOLE K. LEWIS is an assistant professor at Cornell University. Dr. Lewis is also deputy director of the Carl Sagan Institute. She probes exoplanet atmospheres using a combination of observational and theoretical techniques. Dr. Lewis is involved with a number of ground- and space-based observational campaigns aimed at characterizing exoplanet atmospheres. Dr. Lewis was a Sagan Postdoctoral Fellow at the Massachusetts Institute of Technology and the James Webb Space Telescope Project Scientist at the Space Telescope Science Institute before arriving at Cornell. She received her B.S. in physics and mechanical engineering from Worcester Polytechnic Institute, her M.A. in astronomy from Boston University, and her Ph.D. in planetary sciences from the University of Arizona.

BRUCE A. MACINTOSH, *see steering committee entry above*.

AMY MAINZER is a professor at the University of Arizona in the Lunar and Planetary Laboratory. Dr. Mainzer previously served as a senior research scientist at the Jet Propulsion Laboratory (JPL) in the astrophysics division. At JPL, she served as the principal investigator for the NEOWISE mission, which is a NASA spacecraft dedicated to observing near-Earth asteroids and comets using a thermal infrared space telescope. As the NEOWISE principal investigator, Dr. Mainzer's research focuses on characterizing the population of asteroids and comets through statistical measurements of their sizes, orbits, albedos, and rotational states. The mission began life as the Wide-Field Infrared Survey Explorer (WISE), and its original purpose was to carry out an all-sky survey at four infrared wavelengths from 3–22 microns. Dr. Mainzer served as the deputy project scientist for the WISE mission; her responsibilities included flowing down top-level science requirements to the WISE payload components, interpreting payload verification test data, and designing the in-orbit checkout procedures. She has received the NASA Exceptional Scientific Achievement medal for her work on near-Earth objects and the NASA Exceptional Achievement medal for her work on NEOWISE. Prior to joining JPL, Dr. Mainzer worked as a systems engineer at the Lockheed Martin Advanced Technology Center in Palo Alto. Dr. Mainzer is also the principal investigator of a NASA Discovery mission proposal, the Near-Earth Object Camera. She is a member of the NASA Planetary Science Subcommittee.

MARK P. SAUNDERS is an independent consultant. Since retiring from NASA in December 2008, Mr. Saunders has been consulting to various NASA offices providing program/project management and systems engineering expertise. This has included support to the Office of Chief Engineer, the Office of Independent Program and Cost Evaluation, the Mars Program and the Science Office for Mission Assessments (at Langley Research Center). Mr. Saunders has participated in the rewriting of NASA's policy on program/project management; advised and supported the Agency's independent program/project review process; and has supported the review of various programs and projects. At NASA headquarters, he served as director of the independent program assessment office, where he was responsible for enabling the independent review of the Agency's programs and projects at life cycle milestones to ensure the highest probability of mission success. At the Office of Space Science, he served as program manager for the Discovery Program. Mr. Saunders received the Presidential Meritorious Rank Award in 2008; Outstanding Performance awards in 1982 and1994–2008; and the NASA Outstanding Leadership Medals in 1998, 2004, 2006. He earned his B.A. in industrial engineering at the Georgia Institute of Technology.

EVGENYA L. SHKOLNIK is an associate professor of astrophysics at Arizona State University in the School of Earth and Space Exploration. Dr. Shkolnik is an expert on exoplanets and stars, including the Sun, and studies stellar activity and star-planet interactions using ground and space telescopes to answer

questions involving stellar evolution, exoplanet magnetic fields, and planet habitability. She is the principal investigator (PI) of the NASA Star-Planet Activity Research CubeSat (SPARCS) mission, and PI of the Hubble Space Telescope (HST)'s Habitable Zones and M Dwarf Activity Across Time (HAZMAT) program. Asteroid Shkolnik (25156) was named for her. Dr. Shkolnik is also a member of the NASA Astrobiology Institute Virtual Planetary Laboratory, and is on several science and technology advisory committees for upcoming space missions. Shkolnik previously was an astronomer at Lowell Observatory, a Carnegie Postdoctoral Fellow in the Department of Terrestrial Magnetism at the Carnegie Institution for Science, and a National Research Council Postdoctoral Fellow at the University of Hawaii, Manoa. She earned her Ph.D. in astrophysics from the University of British Columbia.

GEORGE SONNEBORN was an astrophysicist at NASA's Goddard Space Flight Center, with 37 years of experience in design, development, and operation of space telescopes. Before retiring from the agency in 2018, Dr. Sonneborn was the project scientist for operations for the James Webb Space Telescope. Prior to that, he was the Hubble Space Telescope acting senior project scientist and the project scientist for the Far Ultraviolet Spectroscopic Explorer. His research interests are supernovae, massive stars, and the atomic abundances in the interstellar medium. Dr. Sonneborn is a member of the AAS and IAU. He received a Ph.D. in astronomy from Ohio State University.

C. MEGAN URRY is the Israel Munson Professor of Physics and Astronomy at Yale University. Dr. Urry is also director of the Yale Center for Astronomy and Astrophysics. She previously served as chair of the Physics Department at Yale and in the presidential line of the American Astronomical Society. Prior to moving to Yale, Urry was a senior astronomer at the Space Telescope Science Institute, which runs the Hubble Space Telescope for NASA. Her scientific research focuses on active galaxies, which host accreting supermassive black holes in their centers. Dr. Urry is a fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science, the American Physical Society and American Women in Science and is a member of the National Academy of Sciences. She received an honorary doctorate from Tufts University and was awarded the American Astronomical Society's Annie Jump Cannon and George van Biesbroeck prizes. Dr. Urry received her Ph.D. in physics from the Johns Hopkins University and her B.S. in physics and mathematics from Tufts University.

PANEL ON ELECTROMAGNETIC OBSERVATIONS FROM SPACE 2

STEVEN M. KAHN, *Chair*, is Cassius Lamb Kirk Professor in the Natural Sciences and professor of particle physics and astrophysics at Stanford University. Dr. Kahn is also the director of the Large Synoptic Survey Telescope (LSST) project at AURA and SLAC National Accelerator Laboratory. Dr. Kahn's research interests include the LSST that will enable a wide array of scientific investigations ranging from studies of moving objects in the solar system to the structure and evolution of the universe as a whole. Prior to joining Stanford University, he held numerous positions at Columbia University, including I.I. Rabi Professor of Physics, professor of physics, and assistant professor of physics. In addition, Dr. Kahn was a Center Postdoctoral Research Fellow at the Harvard- Smithsonian Center for Astrophysics. He is currently a member of both the editorial board for the Cambridge Observing Handbooks for Research Astronomers and the editorial board for the Cambridge Contemporary Astrophysics Series at Cambridge University Press. In addition, Dr.Kahn is a member of the external advisory committee in the particle physics division of the department of physics at the University of Oxford, co-chair of the external advisory committee of the Giant Magellan Telescope, and an outside member of the Astro E Science Working Group run by NASA and ISAS. He is affiliated with the AAS High Energy Astrophysics Division, the APS Astrophysics Division, the AAAS, the American Association of University Professors, and the American Academy of Arts and Sciences. Dr. Kahn is the recipient of many awards including fellowships to the American Association for the Advancement of Science, the American Academy of Arts and Sciences, and the American Physical Society. In addition, he

received the Andrew R. Mikelson Prize in Physics. Dr. Kahn received his Ph.D. in physics from the University of California, Berkeley.

LISA BARSOTTI is a principal research scientist with the Massachusetts Institute of Technology in the Kavli Institute for Astrophysics and Space Research, and part of the Laster Interferometer Gravitational-Wave Observatory (LIGO) Laboratory. Dr. Barsotti's research interests include strong gravity and gravitational radiation, gravitational wave detection, and quantum measurements. In particular, she led the upgrade to the Advanced LIGO detectors to use squeezed vacuum states. She is a fellow of the APS, and she has been awarded the 2019 New Horizons in Physics Prize. Dr. Barsotti earned her Ph.D. in applied physics from Pisa University, Italy.

ALLISON BARTO is a senior program manager at Ball Aerospace. Ms. Barto began her career supporting development of the Advanced Camera for Surveys (ACS) and Cosmic Origins Spectrograph (COS) instruments for the Hubble Space Telescope and most recently spent 17 years supporting the design, build, and test of the James Webb Space Telescope (JWST) as both an optical systems engineer and program manager for the Ball Aerospace effort. This included the optical design, delivery of all opto-mechanical mirror components and electronics for JWST, cryogenic instrument radiators, and the wavefront sensing and control algorithms used to phase the telescope on orbit. In addition to program roles, Ms. Barto led the systems engineering team for the NASA In-Space Assembled Telescope study and serves on the Management Advisory Committee for the European Southern Observatory's Extremely Large Telescope. She is actively involved with the International Society for Optics and Photonics (SPIE) where she currently serves as chair of the Symposia Committee and sits on the Strategic Planning Committee. She is recipient of the 2014 Women in Aerospace Achievement Award for her technical contributions to the JWST optical verification program and the 2017 Aviation Week Program Excellence Award for her work on JWST's cryogenic electronics system. She earned her B.S. in physics from Harvey Mudd College.

MICHAEL BAY is president and chief engineer of Bay Engineering Innovations. Mr. Bay has more than 41 years of experience in systems design and space flight systems engineering on over 20 NASA space missions. He has extensive experience in leading system and detailed design; development, manufacturing; testing; verification; mission planning; launch site and mission operations; and anomaly investigation and resolution activities, both for pre-flight testing and in-orbit activities. Bay is a member of the Avionics Technical Discipline Team of the NASA Engineering and Safety Center (NESC) and a participant in the NESC's Systems Engineering Technical Discipline Team. Mr. Bay also led the systems engineering portion of the NESC's Technical Support to the National Highway Traffic Safety Administration (NHTSA) on the Reported Toyota Motor Corporation (TMC) Unintended Acceleration (UA) Investigation. Mr. Bay received a NASA Public Service Medal for Leadership in Systems Engineering, and NASA's Distinguished Public Service Medal. He was the first recipient of the Institute of Electrical and Electronics Engineers' (IEEE) Judith A. Resnik Field Award for "engineering solutions to urgent spacecraft testing problems and for developments in on-orbit servicing." Mr. Bay received his B.S. in computer science from Loyola University, Maryland.

MARTIN ELVIS is a senior astrophysicist at the Center for Astrophysics | Harvard & Smithsonian. Previously, Dr. Elvis was the science data system group leader at the Chandra X-Ray Center. His research interests include near-Earth asteroid detection and properties, and observations and theory of active galactic nuclei. He is a highly cited astrophysicist (over 30,000 peer citations) who has published some 400 refereed papers. He is a fellow of the AAAS, a member of the Aspen Center for Physics, and is past-chair of the Hubble Space Telescope Users' Committee, and of the High Energy Division of the AAS. Asteroid 9283 Martinelvis is named after him. Dr. Elvis earned his Ph.D. in X-ray astronomy from Leicester University.

CHARLES J. HAILEY is the Pupin Professor of Physics and co-director of the Columbia Astrophysics Laboratory at Columbia University. Dr. Hailey's research interests are observational high energy astrophysics and experimental particle astrophysics. He chairs the Galactic Plane Survey Working Group on the Nuclear Spectroscopic Telescope Array (NuSTAR) mission, and coordinates the NuSTAR legacy observations of the Galactic Center and of TeV gamma-ray sources in conjunction with the Very Energetic Radiation Imaging Telescope Array System (VERITAS) and the High-Altitude Water Cherenkov Gamma-Ray Observatory (HAWC). He is the principal investigator of the General Antiparticle Spectrometer (GAPS) experiment, a balloon-based search for dark matter. He is a member of the American Astronomical Society (AAS) and the High Energy Astrophysics Division of the American Physical Society (APS). Dr. Hailey earned his Ph.D. in physics from Columbia University.

CHRYSSA KOUVELIOTOU is a professor of astrophysics and chair of the Department of Physics at George Washington University (GWU). Prior to GWU, Dr. Kouveliotou was a senior technologist for High-Energy Astrophysics at NASA's Marshall Space Flight Center. She earned her Ph.D. in astrophysics from the Technical University of Munich, Germany. Her research interests focus on high-energy astrophysical transients, in particular gamma ray bursts and magnetars (which she discovered in 1998); she has also published papers in X-ray binaries, solar flares and merging galaxy clusters. Dr. Kouveliotou has been a co-investigator of BATSE/CGRO, Fermi/GBM; she is currently an affiliated scientist of Swift and participates in two working groups of ESA's ATHENA mission. She is the recipient of the Descartes Prize, the Rossi and Heinemann Prizes, and the NASA Exceptional Service Medal. She holds two honorary degrees from Sussex University and the University of Amsterdam; she is an APS and AAAS fellow, and an AAS legacy fellow. Dr. Kouveliotou is a member of the National Academy of Sciences, the American Academy of Arts and Sciences, a foreign member of the Royal Dutch Academy of Sciences, and a corresponding member of the Athens Academy, Greece. She is a fellow of the American Association for the Advancement of Science and served on its executive council. In 2015, the Greek Government awarded her the Commander of the Order of the Honor medal, for excellence in science. In 2021, Dr. Kouveliotou shared the Shaw Prize for Astrophysics with Victoria Kaspi. She has chaired the NASA/Astrophysics Division Roadmap of the next three decades. She served in the ExCom of the NAS/Space Studies Board, in the AAS/HEAD as chair, AAS/vice president, and APS/DAP chair. She currently chairs the IAU/USNC and serves in the AURA Board.

CHARLES R. LAWRENCE is the chief scientist for astronomy and physics at the Jet Propulsion Laboratory. Dr. Lawrence is the project scientist for the U.S. Planck mission and deputy project scientist for the Spitzer Space Observatory. His research interests include measurement and analysis of the cosmic microwave background to understand the geometry and content of the Universe, extragalactic radio sources, and gravitational lensing. He is the recipient of two Exceptional Achievement Medals, two Outstanding Leadership Medals, and a Distinguished Public Service Medal, all from NASA, and was part of the Planck team awarded the Gruber Prize in Cosmology. He has been a member of the AAS since 1983. Dr. Lawrence earned his Ph.D. in physics from the Massachusetts Institute of Technology.

S. HARVEY MOSELEY JR. is vice president for engineering at Quantum Circuits, Inc. Dr. Moseley has long experience with complex systems operating at cryogenic temperatures. He was a key member of the Cosmic Background Explorer science and development team, whose leaders Mather and Smoot won the 2006 Nobel Physics prize for its groundbreaking measurements of the early universe. He invented and led the advance of cryogenic X-ray microcalorimeters, which are central to the scientific capability of current and future X-ray astrophysics missions. He led the creation of microshutter arrays that provide multi-object spectroscopy on JWST. Dr. Moseley has received the Joseph Weber prize of the AAS, the George Goddard Prize of the International Society of Optics and Photonics (SPIE), and was conferred the rank of Distinguished Senior Professional by President Obama. He earned his Ph.D. in astronomy and astrophysics from the University of Chicago.

RESHMI MUKHERJEE is the Helen Goodhart Altschul Professor of Physics and Astronomy at Barnard College. Dr. Mukherjee's research interests are in high-energy astrophysics and astroparticle physics. She uses ground-based atmospheric Cherenkov telescopes to study galactic and extragalactic high-energy gamma-ray sources. One of her current projects is VERITAS, a ground-based gamma-ray observatory. Mukherjee's research also involves the development of next-generation telescope instrumentation for the Cherenkov Telescope Array (CTA). She earned her Ph.D. in physics from Columbia University.

LYMAN A. PAGE JR., *see steering committee entry above*.

GORDON J. STACEY is a professor of astronomy and director of undergraduate studies at Cornell University. Dr. Stacey's research interests center on studies of star formation and its interplay with the interstellar medium across cosmic time. These studies have focused on far-infrared and submillimeter wavelength fine-structure and rotational line emission from abundant atoms, ions, and molecules. Current projects include fine-structure line studies of galaxies both locally and at high redshift, with the Field-Imaging Far-Infrared Line Spectrometer for the Stratospheric Observatory for Infrared Astronomy (SOFIA), the Herschel and Spitzer archives, ALMA, and his ZEUS-2 spectrometer on APEX. Dr. Stacey also is collaborating with NASA Goddard Space Flight Center in the construction of the HIRMES spectrometer for SOFIA, which focuses on protoplanetary disk studies and is constructing an imaging spectrometer for use on Cornell's CCAT-prime telescope. His group is also designing and fabricating new Fabry-Perot mirror technologies. He is an AAS and IAU member and has served on numerous national and international review panels. Dr. Stacey earned his Ph.D. in astronomy from Cornell University.

PANEL ON OPTICAL AND INFRARED OBSERVATIONS FROM THE GROUND

TIMOTHY M. HECKMAN, *Chair*, is the inaugural Dr. A. Hermann Pfund Professor in the Department of Physics and Astronomy at Johns Hopkins University. Dr. Heckman is also the director of the Center for Astrophysical Sciences, where he is responsible for promoting and supporting research in astrophysics, nurturing large-scale projects and providing them with an organizational structure, providing a forum and a focus for strategic planning, fostering cooperation between the different elements of the local astrophysics and space science communities, and providing a structured career path for the non-tenure-track research staff. Dr. Heckman's research interests include galaxy evolution, starbursts, black holes, and active galactic nuclei. He is a member of the GALEX Science Team, a builder of the Sloan Digital Sky Survey (SDSS), chair of the Pan-STARRS1 Science Consortium Board, vice chair of the Board of the Association of Universities for Research in Astronomy, and former chair of the Astrophysical Research Consortium (ARC) Board of Governors, during which time ARC established the SDSS. Dr. Heckman is a member of the National Academy of Sciences. He received his Ph.D. in astronomy from the University of Washington.

DAVID A. BEARDEN is a senior strategist in the Innovation Foundry Office of Formulation at the Jet Propulsion Laboratory (JPL). Dr. Bearden leads teams to develop advanced concepts across JPL's mission directorates. He serves on standing review and other review boards for NASA. Dr. Bearden has considerable expertise concerning the issues and potential solutions in balancing benefit, cost, and risk across a broad array of space systems application areas including science missions, human spaceflight, remote sensing, telecommunications, missile defense, launch, and operations. Prior to joining JPL, Dr. Bearden was general manager of the NASA and Civil Space Division at The Aerospace Corporation, where he was responsible for management and technical leadership of the company's support to NASA headquarters and centers as well as civil space agencies. He has served on the board of trustees for the International Space University (ISU). Dr. Bearden was awarded a Ph.D. in aerospace engineering from the University of Southern California.

DAVID CHARBONNEAU is a professor at Harvard University in the Department of Astronomy. Dr. Charbonneau was previously the R.A. Millikan Postdoctoral Scholar in Astronomy at the California Institute of Technology. His research focuses on the detection and characterization of planets orbiting other stars. He measured the first exoplanet transits, and developed the primary methods which astronomers now regularly use to investigate exoplanet atmospheres. He leads the MEarth project, with his team announced the discovery of several of the closest rocky exoplanets, which are amenable to characterization. His focus on low-mass stars as exoplanet targets has led to several discoveries concerning the physical processes by which theses stars maintain magnetic fields, and how they lose angular momentum as they age. He is a co-investigator in the NASA TESS Mission. Dr. Charbonneau is a member of the National Academy of Sciences. He received his Ph.D. in astronomy at Harvard University.

SUVI GEZARI is an associate professor at University of Maryland. Dr. Gezari has previously been an associate research scientist at Johns Hopkins University and a Hubble Fellow at Johns Hopkins University. Her research focus is on time domain astrophysics. She has used the Pan-STARRS1 (PS1) Survey, the Palomar Observatory surveys iPTF and ZTF at optical wavelengths and the Galaxy Evolution Explorer (GALEX) Time Domain Survey at ultraviolet wavelengths, together with follow-up space-based and ground-based observations from across the electromagnetic spectrum, to discover and characterize transients and study their physical properties. Dr. Gezari has appeared on public television, the history channel, and Canadian public radio discussing her research. She received the NSF CAREER award in 2015 for her research on “Probing the Demographics of Supermassive Black Holes with Time-Domain Observations of Tidal Disruption Events.” Dr. Gezari received her Ph.D. in astronomy from Columbia University.

ANDREA M. GHEZ is professor of physics and astronomy at the University of California, Los Angeles (UCLA), in the Division of Physics and Astronomy. Dr. Ghez has previously held the positions of associate and assistant professor of physics and astronomy at UCLA and was the Hubble postdoctoral research fellow at the University of Arizona. Her expertise is related to the development and application of high spatial resolution infrared imaging techniques applied to the questions of the origin and early life of stars and planets, and the distribution and nature of matter at the center of our galaxy. Her work also strives to understand how a black hole gains mass from its surroundings and what can be learned by analogy about the formation and evolution of galaxies and their central black holes. Dr. Ghez is a member of the National Academy of Sciences and has received the Bakerian Medal and the Crafoord Prize. She received her Ph.D. in physics from the California Institute of Technology.

JENNY E. GREENE is a professor of astrophysical sciences at Princeton University. Dr. Greene has previously been an assistant professor of astronomy at the University of Texas, Austin, and a Carnegie-Princeton postdoctoral fellow at Princeton University. Her expertise is related to black hole mass measurements, black hole/galaxy connections, stellar and gaseous kinematics of galactic nuclei, stellar populations in galaxies, and the low surface brightness universe. Dr. Greene has received the Alfred P. Sloan Fellowship and the Bok Prize from the Harvard University Astronomy Department as well as the Annie Jump Cannon Award from AAS. She received her Ph.D. for astronomy from Harvard University.

J. TODD HOEKSEMA is a senior research scientist at Stanford University in the W.W. Hansen Experimental Physics Laboratory. Dr. Hoeksema has previously been a research associate at Stanford’s Center for Space Science and Astrophysics and Heliophysics Discipline Scientist at NASA HQ. His primary scientific interests include physics of the Sun and heliosphere; solar and coronal magnetic fields; space weather; helioseismology; and education and public outreach. Dr. Hoeksema’s experience includes research administration; system and scientific programming; and the design and operation of instruments to measure solar magnetic and velocity fields from ground and space. He is a Calvin College

distinguished alumni and NASA distinguished public service medal recipient. He received his Ph.D. in applied physics from Stanford University.

JACOBUS M. OSCHMANN, *see steering committee entry above*.

RICHARD W. POGGE is professor and vice chair for Instrumentation at the Ohio State University. Dr. Pogge is a co-discoverer of the Narrow Line Seyfert 1 subclass of AGN and did early work on the ionization morphology of active galactic nuclei. In recent years, he led the building and commissioning of OSU's twin multi-object optical spectrographs for the Large Binocular Telescope (MODS1 & MODS2), and has worked on every major instrument project at OSU since 1989. Dr. Pogge's current research is focused on understanding and revising the absolute metallicity calibration of HII regions in nearby and distant galaxies, a topic of crucial importance for understanding the chemical evolution and growth of galaxies over cosmic time, and he continues work on active galactic nuclei and exoplanets. He received his Ph.D. in astronomy and astrophysics from the University of California, Santa Cruz.

MASSIMO ROBBERTO is an AURA Observatory Scientist at the Space Telescope Science Institute. Dr. Robberto is also a research scientist at Johns Hopkins University. He has previously been an astronomer at the European Space Agency and a staff astronomer at the Max Planck Institut für Astronomie in Heidelberg. At STscI, he is the lead of the JWST/NIRCam team. Before working on the JWST/NIRCam, Dr. Robberto was instrument scientist for the infrared channel of the Wide Field Camera 3 on board the Hubble Space Telescope. He is principal investigator of SCORPIO, the Gen4#3 facility instrument at Gemini South, and principal investigator of SAMOS, an AO-fed MOS for SOAR. His main expertise is in the concept, development, and operations of novel astronomical instrumentation. He has asteroid 2008 QE12 Robberto named after him. Dr. Robberto received his Ph.D. in astronomy from the University of Turin, Italy.

NATASCHA M. FÖRSTER SCHREIBER is a senior staff scientist at the Max-Planck-Institut für extraterrestrische Physik (MPE) in Garching, Germany. Dr. Schreiber has previously held positions as research associate at MPE, and postdoctoral researcher at Leiden Observatory, the Netherlands, and at CEA/DSM/DAPNIA/Service d'Astrophysique in Saclay, France. Her expertise is in the field of galaxy formation and evolution, and her current work focuses on galaxy kinematics, structure, stellar populations, and gas content from spatially resolved and integrated properties using observations in the optical to infrared and millimeter regimes. Dr. Schreiber held a Minerva Fellowship of the Max-Planck-Society in 2008–2013, and was awarded the Degree of Doctor of Science honoris causa (Hon. D.Sc.) from the University of Bath, United Kingdom in 2019. She received her Ph.D. in astrophysics from the Ludwig-Maximilians Universität, Munich, and MPE, Germany.

DAVID R. SILVA is Distinguished Professor of Physics and Astronomy and dean of the College of Sciences at the University of Texas at San Antonio (UTSA). Dr. Silva is a former director of the National Science Foundation's National Optical Astronomy Observatory (NOAO, 2008–2019). His scientific research interests are in the general area of stars and stellar systems, especially as tracers for how galaxies formed and evolved over the last 13 billion years. He has extensive experience with the design, development, and operation of astronomical observatories, telescopes, focal-plane instruments, and data systems for the European, North American, and South American research communities.

PANEL ON PARTICLE ASTROPHYSICS AND GRAVITATION

JOHN F. BEACOM, *Co-Chair*, is the Henry L. Cox Professor of Physics and Astronomy as well as an Arts and Sciences Distinguished Professor at the Ohio State University. Dr. Beacom is also the director of the Center for Cosmology and AstroParticle Physics. His research interests focus on the intersections of

the fields of astrophysics, particle physics, and nuclear physics, especially neutrinos. Prior to joining the Ohio State University, Dr. Beacom was a David N. Schramm Fellow of the Theoretical Astrophysics Group at Fermilab, and a Sherman Fairchild Postdoctoral Scholar at Caltech. He is the recipient of numerous recognitions, including being a Fermilab Distinguished Scholar, a Divisional Associate Editor of Physical Review Letters, a Fellow of the American Physical Society, and the winner of two major teaching awards at the Ohio State University. Dr. Beacom received his Ph.D. in physics from the University of Wisconsin.

LAURA CADONATI, *Co-Chair*, is professor of physics at the Georgia Institute of Technology. Formerly, Dr. Cadanoti was an associate professor at the University of Massachusetts, Amherst. Areas of research include gravitational waves and particle astrophysics, with a focus on the detection, characterization, and astrophysical interpretation of short-duration gravitational wave signals that are produced by cataclysmic astrophysical events such as the collisions of black holes and neutron stars, or core collapse supernovae. She is a member and past deputy spokesperson of the Laser Interferometer Gravitational-Wave Observatory (LIGO) Scientific Collaboration, and a past member of the Borexino Solar Neutrino Collaboration. She is a fellow of the American Physical Society (APS), has chaired the APS Division of Gravity, and is a recipient of the National Science Foundation CAREER Award and the Georgia Institute of Technology outstanding faculty research author award. Dr. Cadonati holds a Ph.D. in physics from Princeton University.

DAVID Z. BESSON is a professor of physics at the University of Kansas. Key areas of research have included particle astrophysics using radio detection methods and astrophysical applications of silicon photomultipliers. In particular, Dr. Besson is currently involved in several projects to detect very-high-energy cosmic rays (primarily protons or neutrinos) from either their radio-wave emissions or radar reflections. He is also involved in studies of anomalous charmed baryon correlations with the Belle and Belle-II experiments. Dr. Besson received a Ph.D. in physics from Rutgers University.

GABRIELA A. GONZÁLEZ, *see steering committee entry above*.

JORDAN A. GOODMAN, *see steering committee entry above*.

ELIZABETH A. HAYS is a research astrophysicist and the chief of the Astroparticle Physics Laboratory at NASA Goddard Space Flight Center. Dr. Hays serves as the project scientist for the Fermi Gamma-Ray Space Telescope. Her research focuses on high-energy studies of astrophysical sites of particle acceleration and development of instrumentation for space-based gamma-ray observatories. She has received the Robert H. Goddard Exceptional Scientific Achievement award and is a fellow of the American Physical Society. Dr. Hays received a Ph.D. in physics from the University of Maryland, College Park.

N. JEREMY KASDIN is the assistant dean for engineering programs at the University of San Francisco. He is also the Eugene Higgins Professor of Mechanical and Aerospace Engineering, emeritus, at Princeton University. Previously, Dr. Kasdin was a member of the Princeton faculty for 20 years and held the post of vice dean of the School of Engineering and Applied Science. Prior to that, he was the chief systems engineer for NASA's Gravity Probe B spacecraft. While at Princeton, he studied techniques for high-contrast imaging from ground and space using coronagraphs and starshades. Dr. Kasdin was the principal investigator for the Coronagraphic High Angular Resolution Imaging Spectrograph (CHARIS) instrument on the Subaru Telescope on Maunakea, Hawai'i. He is the adjutant scientist for the coronagraph instrument on NASA's Wide Field Infrared Survey Telescope. He earned his Ph.D. in aeronautics and astronautics from Stanford University.

DAVID B. KIEDA is a professor at the University of Utah (UU) in the Department of Physics and Astronomy. He also serves as the dean of the UU Graduate School. Dr. Kieda is the head of the UU experimental gamma-ray astronomy research group. He has led the development of new technologies for observational high-energy astrophysics, including work on the Fly's Eye/High-Resolution Fly's Eye, the Very Energetic Radiation Imaging Telescope Array System (VERITAS), HAWC, and the Cherenkov Telescope Array (CTA) observatories. Dr. Kieda also works on the development of techniques for visible band imaging of nearby hot stars with an angular resolution better than 100 micro-arc seconds. He received the Utah Governor's Medal of Science and Technology and is a fellow of the American Physical Society. He earned his Ph.D. in physics from the University of Pennsylvania.

ANDREA N. LOMMEN is a professor and the chair of the Physics and Astronomy Department at Haverford College. Previously, she held the same positions at Franklin and Marshall College. Dr. Lommen has founded efforts in gravitational wave detection using pulsars through both the North American Nanohertz Observatory of Gravitational Waves and the International Pulsar Timing Array. She is currently leading efforts to demonstrate pulsar timing capabilities in the x-ray regime as part of NASA's Neutron Star Interior Composition Explorer. Dr. Lommen has received a National Science Foundation CAREER award. She received a Ph.D. in astrophysics from the University of California, Berkeley.

BRIAN D. METZGER is a professor at Columbia University in the Department of Physics. His research covers a wide range of topics in theoretical high-energy astrophysics, mostly related to compact objects, nucleosynthesis (astrophysical origin of the elements), and the electromagnetic counterparts of gravitational wave sources. Dr. Metzger has received a New Horizons Breakthrough Prize in Physics and a Bruno Rossi Prize of the American Astronomical Society. He earned his Ph.D. in physics from the University of California, Berkeley.

JAMES H. YECK is a researcher with the University of Wisconsin, Madison. Mr. Yeck serves as the interim project director for the Cosmic Microwave Background-Stage 4 (CMB-S4) project. Previously, he was the director general of the European Spallation Source (ESS) in Lund, Sweden, and the project director of the IceCube South Pole Neutrino Observatory. Mr. Yeck has more than 30 years of project director and project manager experience leading projects in both federal and contractor roles. He currently chairs and serves as a member of numerous advisory committees for projects and facilities sponsored by the Department of Energy and National Science Foundation, including LIGO and the Large Synoptic Survey Telescope. He holds an M.S. in mechanical and nuclear engineering from Northwestern University.

NICOLAS YUNES is a professor of physics at the University of Illinois, Urbana-Champaign. Previously, Dr. Yunes was an associate professor of physics and one of the founding directors of the eXtreme Gravity Institute at Montana State University. Key areas of research have included gravitational wave theory, modeling and data analysis with ground- and space-based detectors, black hole and neutron star theory, and tests of general relativity with gravitational waves, binary pulsars, and solar system observations. He has received the Young Scientist Prize of the International Union of Pure and Applied Physics and the International Society on General Relativity and Gravitation, the NASA Einstein Fellowship, and the Juergen Ehlers Thesis Prize from the International Society on General Relativity and Gravitation. Dr. Yunes is an editor of *Classical and Quantum Gravity*. He received a Ph.D. in physics from Pennsylvania State University.

PANEL ON RADIO, MILLIMETER, AND SUBMILLIMETER OBSERVATIONS FROM THE GROUND

ANDREW J. BAKER, *Chair,* is a professor of physics and astronomy at Rutgers, The State University of New Jersey. Dr. Baker's research interests focus on the use of radio, millimeter, and submillimeter wavelength observations of interstellar matter to probe galaxy evolution in the nearby and distant universe. Prior to joining Rutgers, he worked at the University of Maryland as a National Radio Astronomy Observatory (NRAO) Jansky Fellow and at the Max-Planck-Institut für extraterrestrische Physik as a postdoctoral researcher with the infrared/submillimeter astronomy group. Dr. Baker is a fellow of the American Association for the Advancement of Science, a former Fulbright scholar, a former Defense Science Study Group member, and a recipient of the Warren I. Susman Award for Excellence in Teaching. He received his Ph.D. in astronomy from the California Institute of Technology.

HÉCTOR G. ARCE is a professor of astronomy at Yale University. Dr. Arce's research interests include star formation; feedback from young stellar objects, molecular clouds, and cores; and the physical and chemical processes in the interstellar medium. To conduct his research, he mostly uses radio, millimeter, and sub-millimeter telescopes. Prior to joining Yale University, Dr. Arce was an NSF Astronomy and Astrophysics postdoctoral fellow at the American Museum of Natural History in New York, and a postdoctoral researcher in the Owens Valley Radio Observatory millimeter array group at the California Institute of Technology in Pasadena. He has served in several radio/millimeter/sub-millimeter proposal review committees and in the Arecibo Observatory users and scientific advisory committee. Dr. Arce is the recipient of a National Science Foundation CAREER award. He received his Ph.D. in astronomy from Harvard University.

RAVINDER S. BHATIA is associate project manager for the Thirty Meter Telescope (TMT). Dr. Bhatia has worked on international collaborations in technology development for more than 25 years, in astronomy, Earth observation, and oceanography. Previously, he was project systems engineer for the Atacama Large Millimeter/submillimeter Array (ALMA) in Chile. As senior thermal/cryogenics engineer at the European Space Agency, Dr. Bhatia supported the development of the Planck Space Telescope and the MIRI camera for the James Webb Space Telescope, as well as serving as technical officer for technology research and development contracts with industry, government research facilities, and academia. He was visiting research fellow at the UK National Oceanography Centre. As senior postdoctoral scholar at Caltech's Division of Physics, Mathematics and Astronomy, his research focused on developing instruments to measure the Cosmic Microwave Background. He worked in industry as an Aeronautical Engineer for Lucas Aerospace. Dr. Bhatia is a senior member of the American Institute of Aeronautics and Astronautics. Ravinder Bhatia received his Ph.D. in experimental astrophysics and aerospace Engineering from Queen Mary College.

TRACY E. CLARKE is a research astronomer at the U.S. Naval Research Laboratory (NRL) in the Remote Sensing Division. Dr. Clarke's primary research interests involve understanding the large-scale diffuse emission in clusters of galaxies, and their relation to the mergers of clusters of galaxies and to the injection of energy by the huge relativistic jets produced episodically by the supermassive black holes at the centers of galaxies. She uses both X-ray and radio astronomy in her research. She also has made important contributions in the development of radio astronomy hardware. As the current VLA Low-Band Ionosphere and Transient Experiment (VLITE) Project Scientist and the System Scientist for the Long Wavelength Array from 2011–2017, she has a prominent role in advancing the state of the art in synthesis imaging and instrumentation at low radio frequencies. Dr. Clarke is a former Jansky Fellow at the National Radio Astronomy Observatory. She is a member of the American Astronomical Society, the International Union of Radio Science, and the International Astronomical Union, and served on the SKA Organization's Science and Engineering Advisory Committee for the Square Kilometre Array from 2017 to 2021. Dr. Clarke holds a Ph.D. in astronomy from the University of Toronto.

MATT A. DOBBS is a professor at McGill University (Canada) in the Department of Physics. He is also an associate member of the Department of Electrical and Computer Engineering. Dr. Dobbs is a senior

fellow in the Canadian Institute for Advanced Research Gravity and the Extreme Universe program, and a member of the Royal Society of Canada College of New Scholars. His research group at McGill specializes in the development of novel instrumentation and experiments to explore the universe with millimeter wavelength observations of the cosmic microwave background (CMB) radiation and radio observations of 21 cm hydrogen emission and fast radio transients. Dr. Dobbs is the recipient of the 2019 Killam Research Fellowship in Natural Sciences for his project, titled “Unveiling the Cosmos with a New Paradigm Digital Radio Telescope,” involving the recently developed Canadian Hydrogen Intensity Mapping Experiment, (CHIME). He was awarded the inaugural Owen Chamberlain Fellowship at the Lawrence Berkeley Laboratory (U.S.) and earned a Sloan Fellowship. Dr. Dobbs was named Canada Research Chair (T2) in Astro-particle Physics for two terms from 2006 to 2015. He was awarded the inaugural Dunlap Award for Innovation in Astronomical Research Tools and the Canadian Association of Physicists Herzberg Medal. He earned his Ph.D. in experimental particle physics from the University of Victoria (Canada).

DAVID L. KAPLAN is an associate professor of physics at the University of Wisconsin, Milwaukee. Dr. Kaplan’s primary research interests as a multi-wavelength astronomer include compact objects (white dwarfs, neutron stars, and black holes), as well as multi-wavelength and multi-messenger transients. Prior to joining the University of Wisconsin, Milwaukee, he was a Hubble Fellow at the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara, and a Pappalardo Fellow and Hubble Fellow at the Massachusetts Institute of Technology. Dr. Kaplan is co-PI of the Australian Square Kilometre Array Pathfinder (ASKAP) Variables and Slow Transients (VAST) Survey Project, and is a member of the American Astronomical Society, the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), the Zwicky Transient Facility (ZTF), and the Murchison Widefield Array (MWA) collaborations. He serves on the editorial board of *Monthly Notices of the Royal Astronomical Society*. He received his Ph.D. in astrophysics from California Institute of Technology.

DANIEL P. MARRONE is an associate professor of astronomy at the University of Arizona. He is also an associate astronomer at Steward Observatory. Dr. Marrone was previously a Hubble Fellow and a Jansky Fellow at the Kavli Institute for Cosmological Physics at the University of Chicago. His research addresses a number of topics, including the physics of black holes, the formation of early galaxies, and cosmology. His work often relies on the construction of new instruments, primarily at centimeter to submillimeter wavelengths. He is chair of the Event Horizon Telescope (EHT) collaboration science council and a member of the South Pole Telescope collaboration. Dr. Marrone served on the Atacama Large Millimeter/Submillimeter Array (ALMA) Science Advisory Committee from 2018 to 2020 and the National Radio Astronomy Observatory (NRAO) Users Committee from 2014 to 2020. He is the recipient of a National Science Foundation CAREER Award and, as a member of the EHT collaboration, the NSF Diamond Achievement Award and the Breakthrough Prize in Fundamental Physics. Dr. Marrone earned a Ph.D. in astronomy from Harvard University.

LYNN D. MATTHEWS is a research scientist at the Massachusetts Institute of Technology Haystack Observatory. Dr. Matthews specializes in radio wavelength studies of evolved stars and on the deployment of new technologies for observational radio astronomy. She is part of the Event Horizon Telescope team that used the technique of very long baseline interferometry (VLBI) to achieve the first ever image of a supermassive black hole. She served as commissioning scientist for the Atacama Large Millimeter/Submillimeter Array (ALMA) Phasing Project that brought millimeter VLBI capabilities to ALMA, and is currently principal investigator of the ALMA Phasing Project Phase 2. Previously, she held appointments as a Jansky Fellow at the National Radio Astronomy Observatory and as a Clay Fellow at the Harvard-Smithsonian Center for Astrophysics. Dr. Matthews is a member of the American Astronomical Society and the International Astronomical Union. She received a Ph.D. in astronomy from the State University of New York, Stony Brook.

JOAN R. NAJITA is an astronomer and chief scientist at NSF's NOIRLab, where she has been a scientific staff member for the past 22 years. As chief scientist, Dr. Najita is responsible for science planning, science communications, and the health of the scientific environment at NOIRLab. Her research interests include star and planet formation; low mass stars and brown dwarfs; the Milky Way; infrared spectroscopy; massively multiplexed wide-field spectroscopy; and science sociology and resource allocation practices in astronomy. She is also interested in the future of science publications, communicating science to the public, and the role of science in society. A recipient of the Annie Jump Cannon Award in Astronomy, Dr. Najita is a fellow of the John Simon Guggenheim Memorial Foundation, a fellow of the American Association for the Advancement of Science, and a member of the Aspen Center for Physics and the American Astronomical Society. She received her Ph.D. from the University of California, Berkeley.

RICHARD L. PLAMBECK is a research astronomer at the University of California, Berkeley. Dr. Plambeck's research focuses on the development of instrumentation for millimeter wavelength astronomy, and on high-resolution observations of star-forming regions. He helped construct the receivers used on the Berkeley-Illinois-Maryland Array (BIMA) and the Combined Array for Research in Millimeter Astronomy (CARMA). The success of those telescopes led ultimately to the construction of ALMA, the world's premier telescope at mm wavelengths. Dr. Plambeck served on numerous ALMA design reviews and on the ALMA Science Advisory Committee. He received a Ph.D. in physics from the University of California, Berkeley.

JEAN L. TURNER, *see steering committee entry above*.

PANEL ON STATE OF THE PROFESSION AND SOCIETAL IMPACTS

MARGARET M. HANSON, *Co-Chair*, is divisional dean and a professor in the Department of Physics at the University of Cincinnati. Dr. Hanson's research interests include studying massive stars and stellar clusters, along with imaging simulations of stellar clusters that better constrain the properties of resolved clusters. Dr. Hanson was the associate editor-in-chief of the *Astronomical Journal* for 8 years. Prior to joining her current organization, she was a Hubble Postdoctoral fellow at the University of Arizona. She is the recipient of numerous awards and honors, including the Leading Women of Cincinnati Science and Technology Award, the Edith C. Alexander Award for Distinguished Teaching, the National Science Foundation CAREER award, and the Sigma Xi Young Investigator Award. Dr. Hanson is a member of the Astronomical Society of the Pacific and the American Astronomical Society. She received her Ph.D. in astrophysics from the University of Colorado.

ENRICO RAMIREZ-RUIZ, *Co-Chair*, is a professor and the Vera Rubin Chair of Astronomy and Astrophysics at the University of California, Santa Cruz. Dr. Ramirez-Ruiz is also the director of the Lamat Institute, where he works vigorously to support the promotion and retention of women and historically marginalized students in science, technology, engineering, and mathematics (STEM). Dr. Ramirez-Ruiz's research interests include high-energy astrophysics and gravitational wave astronomy. Prior to joining the University of California, Santa Cruz, he was a John Bahcall Fellow at the Institute for Advanced Study at Princeton University. Dr. Ramirez-Ruiz is the recipient of numerous awards and honors, including a Packard Fellowship for Science and Engineering, the National Science Foundation CAREER Award, the Kavli Fellowship from the National Academy of Science, the Radcliffe Fellowship from Harvard University, the Niels Bohr Professorship from the Danish National Research Foundation, the Edward A. Bouchet Award from the American Physical Society, the HEAD Mid-Career Prize from the American Astronomical Society, and a fellowship from the American Academy of Arts and Sciences. Dr. Ramirez-Ruiz received his Ph.D. in astronomy and astrophysics from the University of Cambridge.

GURTINA BESLA is an assistant professor of astronomy at the University of Arizona. Dr. Besla is also an assistant astronomer at Steward Observatory. Her research interests focus on understanding the formation and evolution of low-mass dwarf galaxies, particularly the Large and Small Magellanic Clouds, through numerical simulations. Dr. Besla is the principal investigator of the University of Arizona's TIMESTEP program, which is focused on increasing the presence of underrepresented minority students in the physical sciences. She earned her Ph.D. in astronomy from Harvard University.

PATRICIA T. BOYD is an astrophysicist at NASA's Goddard Space Flight Center. Dr. Boyd serves as the chief of the Exoplanets and Stellar Astrophysics Laboratory and as project scientist for the Transiting Exoplanet Survey Satellite (TESS) mission at NASA's Goddard Space Flight Center. Her research interests focus on the long-term variability in stellar binaries, star-planet interaction, and accretion onto stellar-mass and supermassive compact objects. Dr. Boyd has led the Guest Investigator Programs for several operating NASA missions, including the Rossi X-Ray Timing Explorer, Swift, and TESS. She spent 2 years at NASA headquarters, where she served as the program scientist for the Kepler mission through launch, commissioning, and early operations, while also serving as the GALEX program scientist and managing the Origins of Solar Systems grants portfolio for the Astrophysics Division. Dr. Boyd is a member of the NASA Astrophysics Advisory Committee (formerly the NASA Astrophysics Subcommittee). She was the Goddard lead for the National Astronomy Consortium, an internship program focused on recruiting and retaining STEM professionals from underserved populations, and co-organized the Women in Astronomy Roundtable at Goddard College. Dr. Boyd is also co-creator of the AstroCappella project, a musical exploration of the universe used in classrooms and in live performances. She has been recognized by NASA for her work several times, including exceptional achievement for diversity and EEO and exceptional outreach achievement for the Hubble 25th anniversary. Dr. Boyd earned her Ph.D. in physics and atmospheric science from Drexel University.

KATHRYNE J. DANIEL is assistant professor of physics at Bryn Mawr College. Dr. Daniel's research interests are in galaxy evolution and dynamics. She is a member of the Society for Advancing Chicanos/Hispanics and Native Americans in Science, serves on the AAS Committee on the Status of Minorities in Astronomy, and is a committee member for the Division for Dynamical Astronomy. Dr. Daniel also co-organized a workshop at the AAS on combating racism in astronomy. She received the American Dissertation Fellowship from the American Association of University Women for both her academic work and her role in promoting women in astrophysics. Dr. Daniel earned her Ph.D. in physics and astronomy from Johns Hopkins University.

MARTHA P. HAYNES is the Distinguished Professor of Arts and Sciences in Astronomy at Cornell University. Dr. Haynes's research interests focus on observational cosmology, galaxy evolution, and techniques of radio astronomy. She is a fellow of the American Association for the Advancement of Science and has been elected to the American Academy of Arts and Sciences. Dr. Haynes has received the Henry Draper Medal for investigations in astronomical physics for her work in mapping the distribution of galaxies in the universe. She has been recognized at Cornell for her commitment to undergraduate education and mentoring. Dr. Haynes earned her Ph.D. in astronomy at Indiana University. She has previously served on the Board of Physics and Astronomy, the Division Committee on Engineering and Physical Sciences, and the Report Review Committee at the National Academies, and was co-vice chair of the 2010 decadal survey.

JEDIDAH C. ISLER is an assistant professor of physics and astronomy at Dartmouth University and a consultant and speaker. Dr. Isler's research interests focus on studying blazars using multi-wavelength observations of their particle jets. She is a well-known speaker and advocate for women of color in science. Dr. Isler founded Vanguard: Conversations with Women of Color in STEM, a panel discussing the experiences of women of color in STEM. She also founded and leads the STEM en Route to Change Foundation with the goal to use STEM as a tool for social justice. Dr. Isler received the American

Astronomical Society Roger Doxsey Dissertation Prize and became a TED Fellow. She earned her Ph.D. in astrophysics from Yale University.

RACHEL L. IVIE is the senior director of education and research at the American Institute of Physics (AIP). In this capacity, Dr. Ivie is responsible for the Center for the History of Physics, the Niels Bohr Library and Archives, the Society of Physics Students, and the Statistical Research Center. Prior to her appointment as senior director, Dr. Ivie served AIP in the Statistical Research Center for 21 years, both as assistant and associate director, before leading it entirely as director. Her research interests include physics and astronomy faculty in 4-year institutions, women and underrepresented groups in physics and astronomy, and employment and career paths in physics and astronomy. Dr. Ivie has carried out a number of studies related to the career outlook of women in physics, including on tenure and promotion practices for male and female faculty. She also completed an NSF-funded longitudinal study on gender differences in career outcomes for astronomy graduate students. Dr. Ivie earned her Ph.D. in sociology at the University of North Carolina, Chapel Hill.

KATHRYN V. JOHNSTON is a professor of astronomy at Columbia University. Past appointments include postdoctoral membership in the Institute for Advanced Study, junior faculty at Wesleyan University, and 3 years as the chair of the Columbia Astronomy Department. Dr. Johnston's research interests focus on the dynamics, formation, interactions, and evolution of the galaxy, stellar populations, and the Milky Way and Local Group. Beyond her own research, she is committed to enabling science through community projects and networks. At Columbia, Dr. Johnston helped move the institution to a shared model for research computing and is currently the chair of the Committee for Equity and Diversity in the School of Arts and Sciences. She has also led discussions on women in science at more than 20 departments nationwide in the past decade. Dr. Johnston is just starting a joint appointment as Dynamics Group Leader at the Flatiron Institute's Center for Computational Astrophysics. She earned her Ph.D. in astronomy from the University of California, Santa Cruz.

CASEY W. MILLER is the associate dean for Research and Faculty Affairs for the College of Science at the Rochester Institute of Technology. Dr. Miller is also a professor in the School of Chemistry and Material Science. His research interests include experimental, nanoscale magnetic materials, and he is a nationally recognized expert in STEM graduate education. Dr. Miller serves as director of the National Science Foundation (NSF) INCLUDES Alliance: IGEN's Inclusive Practices Hub. He has served as the director of the American Physical Society (APS) Bridge Program's site at the University of South Florida, which strives to increase the number of physics Ph.D.s awarded to underrepresented minority students. Dr. Miller has also served on the APS Committee on Minorities, and he was the chair of the 2017 APS Graduate Education and Bridge Program Conference. He earned his Ph.D. in physics from the University of Texas, Austin.

JESÚS PANDO is an associate professor of physics and astrophysics at DePaul University and currently serves as the chair of the department. Dr. Pando's research interests focus on uncovering structure in a noisy environment, such as large-scale structure formation in the universe. He is a member of the Society for the Advancement of Chicanos/Hispanics and Native Americans in Science and a board member of the National Society of Hispanic Physicists, both with the goal of dealing with issues faced by underrepresented students and professionals in STEM. Dr. Pando earned his Ph.D. in physics from the University of Arizona.

JULIE R. POSSELT is an associate professor of higher education at the University of Southern California at the Rossier School of Education. Dr. Posselt's research examines institutionalized inequalities in higher education and methods to reduce inequities and encourage diversity. She has written three books focusing on equity and inclusion in higher education, as well as numerous articles and papers on the subject. Dr. Posselt completed the National Academy of Education's first national study of graduate student mental

health, and directs the National Science Foundation-funded California Consortium for Inclusive Doctoral Education and the Inclusive Graduate Education Research Hub. She is associate editor of the *Journal of Higher Education*. Dr. Posselt earned her Ph.D. in higher education from the University of Michigan.

JANE R. RIGBY is an astrophysicist at the NASA Goddard Space Flight Center. Dr. Rigby served for 9 years as a project scientist for the James Webb Space Telescope (JWST) and is currently the operations project scientist. She was a NASA headquarters-appointed member of the Science and Technology Definition Team for the NASA Large UV/Optical/IR Surveyor (LUVOIR) mission concept. Dr. Rigby's research interests focus on observations of star-forming galaxies, supermassive black holes, and gravitational lensing as a tool to study galaxies. She has received numerous awards, including the John C. Lindsay Memorial Award for Space Science, the Robert H. Goddard Award for Exceptional Achievement for Science, and the Robert H. Goddard Award for Diversity and Equal Employment Opportunity. Dr. Rigby co-organized the "Inclusive Astronomy 2015" conference, served as a founding member of the AAS's Working Group on LGBTIQ Equality, and served on the AAS Committee for Sexual-Orientation and Gender Minorities in Astronomy. She has given public talks to large audiences including TEDx, the Library of Congress, and two conferences for undergraduate women in physics, and has lectured on the impact of gay activist and astronomer Frank Kameny. Dr. Rigby earned her Ph.D. in astronomy from the University of Arizona.

WILLIE S. ROCKWARD is chair and professor of physics at Morgan State University. Dr. Rockward's research interests include micro/nano optics lithography, extreme ultraviolet interferometry, metamaterials, and the spectroscopy of binary stars. He is the currently the president of the National Society of Black Physicists. As chair of his department, Dr. Rockward investigated the barriers faced by the physics departments of Historically Black Colleges and Universities (HBCU) and launched the "We C.A.R.E." approach meant to improve the overall number of African American physicists. He gave the keynote speech at the Conference for Underrepresented Minority Physicists in 2017. Dr. Rockward received his Ph.D. in physics from Georgia Institute of Technology.

KEIVAN G. STASSUN, *see steering committee entry above*.